emergence of regions of akinesis or dyskinesis not present at rest. Stress echocardiography, like stress myocardial perfusion imaging, is more sensitive than exercise electrocardiography in the diagnosis ofIHD. The relative advantages of stress echocardiography and stress radionuclide perfusion imaging in the diagnosis of IHD are shown inTable 244-1.

Echocardiography or radionuclide angiography should be carried out to assess left ventricular function in patients with chronic stable angina and in patients with a history of a prior myocardial infarction, pathologic Q waves, or clinical evidence of heart failure.

Coronary Arteriography (See also Chap. 228) This diagnostic method outlines the coronary anatomy and can be used to detect important evidence of coronary atherosclerosis or to exclude this condition. By this means, one can assess the severity of obstructive lesions and, when coronary arteriography is combined with left ventricular angiocardiography, can evaluate both global and regional function of the left ventricle.

Indications Coronary arteriography is indicated in (1) patients with chronic stable angina pectoris who are severely symptomatic despite medical therapy and who are being considered for revascularization, i.e., a percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG); (2) patients with troublesome symptoms that present diagnostic difficulties in whom there is need to confirm or rule out the diagnosis ofIHD; (3) patients with known or possible angina pectoris who have survived sudden cardiac death; and (4) patients judged to be at high risk of sustaining coronary events based on signs of severe ischemia on noninvasive testing, regardless of the presence or severity of symptoms (see below).

Examples of other clinical situations include:

1. Patients with chest discomfort suggestive of angina pectoris but a negative or nondiagnostic stress test who require a definitive diagnosis for guiding medical management, alleviating psychological stress, career or family planning, or insurance purposes.

2. Patients who have been admitted repeatedly to the hospital for suspected acute myocardial infarction but in whom this diagnosis has not been established and in whom the presence or absence ofCADshould be determined.

3. Patients with careers that involve the safety of others (e.g., airline pilots) who have questionable symptoms, suspicious or positive noninvasive tests, and in whom there are reasonable doubts about the state of the coronary arteries.

4. Patients with aortic stenosis or hypertrophic cardiomyopathy and angina in whom the chest pain could be due toIHD.

5. Male patients aged 45 and females aged 55 years of age or older who are to undergo a cardiac operation, such as valve replacement or repair and who may or may not have clinical evidence of myocardial ischemia.

6. Patients who are at high risk after myocardial infarction because of the recurrence of

angina or the presence of heart failure, frequent ventricular premature contractions, or signs of ischemia in the stress test.

7. Patients with angina pectoris, regardless of severity, in whom noninvasive testing indicates a high risk of coronary events.

8. Patients in whom coronary spasm or another nonatherosclerotic cause of myocardial ischemia (e.g., coronary artery anomaly, Kawasaki's disease) is suspected.

PROGNOSIS

The principal prognostic indicators in patients withIHD are the functional state of the left ventricle, the location and severity of coronary artery narrowing, and the severity or activity of myocardial ischemia. Angina pectoris of recent onset, unstable angina, angina that is unresponsive or poorly responsive to medical therapy or is accompanied by symptoms of congestive heart failure all indicate an increased risk for adverse coronary events. The same is true for the physical signs of heart failure, episodes of pulmonary edema, transient third heart sounds, or mitral regurgitation or for echocardiographic (or roentgenographic) evidence of cardiac enlargement. An abnormal restingECG or positive evidence of myocardial ischemia during a stress test also indicates increased risk. Most importantly, the following signs during noninvasive testing indicate a high risk for coronary events: a strongly positive exercise test showing onset of myocardial ischemia at low workloads [³0.1 mV ST-segment depression before completion of stage II (Bruce protocol) of the exercise test;³0.2 mV ST depression in any stage; ST depression for >5 min following the cessation of exercise; a decline in systolic pressure >10 mmHg during exercise; the development of ventricular tachyarrhythmias during exercise]; the development of large or multiple perfusion defects or increased lung uptake during stress radioisotope perfusion imaging; and a decrease in left ventricular ejection fraction during exercise on radionuclide ventriculography or during stress echocardiography. Conversely, patients who can complete stage III of the Bruce exercise protocol and have a normal stress perfusion scan or negative stress echocardiographic evaluation are at very low risk of future coronary events.

On cardiac catheterization, elevations in left ventricular end-diastolic pressure and ventricular volume and a reduced ejection fraction are the most important signs of left ventricular dysfunction and are associated with a poor prognosis. Patients with chest discomfort but normal left ventricular function and normal coronary arteries have an excellent prognosis. In patients with normal left ventricular function and mild angina but with critical stenoses (³70% luminal diameter) of one, two, or three epicardial coronary arteries, the 5-year mortality rates are approximately 2, 8, and 11 percent, respectively. Obstructive lesions of the left anterior descending coronary artery proximal to the origin of the first septal artery are associated with a greater risk than are lesions of the right or left circumflex coronary artery, since the former vessel usually perfuses a greater quantity of myocardium. Stenosis (>50% luminal diameter) of the left main coronary artery is associated with a mortality rate of about 15% per year. The segmental atherosclerotic plaques in epicardial arteries go through phases of inflammatory cellular activity, degeneration, endothelial instability, abnormal vasomotion, platelet aggregation, and fissuring or hemorrhage. These factors can temporarily worsen the stenosis and cause abnormal reactivity of the vessel wall, thus exacerbating the manifestations of

ischemia. The recent onset of symptoms, the appearance of severe ischemia during stress testing, and unstable angina pectoris (p. 1508) all reflect episodes of rapid progression in coronary lesions.

With any degree of obstructiveCAD, mortality is greatly increased when left ventricular function is impaired; conversely, at any level of left ventricular function, the prognosis is influenced importantly by the quantity of myocardium perfused by the critically obstructed vessels. Therefore, it is useful to collect all the evidence substantiating past myocardial damage (ECG and ventriculographic evidence of myocardial infarction), residual left ventricular function (ejection fraction and wall motion), and risk of future damage from coronary events (extent of coronary disease and severity of ischemia defined by noninvasive stress testing). The larger the amount of established myocardial necrosis, the less the heart is able to withstand additional damage and the poorer the prognosis. All the above signs of past damage plus the risk of future damage should be considered indicators of risk.

TREATMENT

Each patient must be evaluated individually with respect to his or her expectations and goals, control of symptoms, and prevention of adverse clinical outcomes such as myocardial infarction and premature death. The degree of disability as well as the physical and emotional stress that precipitate angina must be carefully recorded in order to set treatment goals. Each management plan should consist of the following: (1) explanation and reassurance, (2) identification and treatment of aggravating conditions, (3) adaptation of activity, (4) treatment of risk factors that will decrease the occurrence of adverse coronary outcomes, (5) drug therapy for angina, and (6) consideration of mechanical revascularization.

Explanation and Reassurance Patients withIHD need to understand their condition as best they can and to realize that a long and useful life is possible even though they suffer from angina pectoris or have experienced and recovered from an acute myocardial infarction. Offering case histories of persons in public life who have lived with coronary disease as well as results of national studies showing improved outcomes can be of great value when encouraging patients to resume or maintain activity and return to their occupation. A planned program of rehabilitation can encourage patients to lose weight, improve exercise tolerance, and control risk factors with more confidence.

Identification and Treatment of Aggravating Conditions A number of conditions may either increase oxygen demand or decrease oxygen supply to the myocardium and may precipitate or exacerbate angina. Aortic valve disease and hypertrophic cardiomyopathy may cause angina and should be excluded or treated. Obesity, hypertension, and hyperthyroidism may be managed successfully in order to reduce the frequency of anginal attacks. Decreased myocardial oxygen supply may be due to reduced oxygenation of the blood (e.g., in pulmonary disease or, when carboxyhemoglobin is present, due to cigarette or cigar smoking) or decreased oxygen-carrying capacity (e.g., in anemia). Correction of these abnormalities, if present, may reduce or even eliminate angina pectoris.

Adaptation of Activity Therapy of angina due to episodes of myocardial ischemia

consists of eliminating the discrepancy between the demand of the heart muscle for oxygen and the ability of the coronary circulation to meet this demand. Most patients can be made to understand this fundamental concept and utilize it in the rational programming of activity. Many tasks that ordinarily evoke angina may be accomplished without symptoms simply by reducing the speed at which they are performed. Patients must appreciate the diurnal variation in their tolerance of certain activities and should reduce their energy requirements in the morning and immediately after meals. Sometimes it is helpful to alter the eating pattern, taking small and more frequent meals.

It may be necessary to recommend a change in employment or residence to avoid physical stress; however, with the exception of manual laborers, most patients withIHD can continue to function merely by allowing more time to complete each task. In some patients, anger and frustration may be the most important factors precipitating myocardial ischemia. If these cannot be avoided, training in stress management may be useful. A treadmill exercise test to determine the approximate heart rate at which ischemicECGchanges or symptoms develop may be helpful in the development of a specific exercise program.

Physical conditioning usually improves the exercise tolerance of patients with angina and exerts substantial psychological benefits. It may also improve the chances of surviving a myocardial infarction. An exercise program within the limits of each patient's threshold for the development of angina pectoris should be encouraged.

Treatment of Risk Factors Although the treatment of risk factors was developed for the primary prevention of coronary atherosclerosis, there is growing evidence that it can reduce the occurrence of angina, myocardial infarction, and death both in subjects without provenIHD as well as in those with a history of chronic angina or an acute coronary syndrome. A *family history* of premature IHD is an important indicator of increased risk and should trigger a search for treatable risk factors such as hyperlipidemia, hypertension, and diabetes. *Obesity* impairs the treatment of other risk factors and increases the risk of adverse coronary events. In addition, obesity is often accompanied by two other risk factors -- hypertension and hyperlipidemia. The treatment of obesity and these accompanying risk factors is an important component of any management plan.

Cigarette smoking accelerates coronary atherosclerosis in both sexes and at all ages and increases the risk of myocardial infarction and death. By increasing myocardial oxygen needs and reducing oxygen supply it aggravates angina. Smoking cessation studies have demonstrated important benefits with a significant decline in the occurrence of these adverse outcomes. The physician's message must be clear and strong and supported by programs that achieve and monitor abstinence (Chap. 390). *Hypertension* (Chaps. 35 and246) is associated with increased risk of adverse clinical events from coronary atherosclerosis as well as stroke. In addition, the left ventricular hypertrophy that results from sustained hypertension aggravates ischemia. There is evidence that long-term, effective treatment of hypertension can decrease the occurrence of adverse coronary events. *Diabetes mellitus* (Chap. 333) accelerates coronary and peripheral atherosclerosis and is frequently associated with dyslipidemias and increases in the risk of angina, myocardial infarction, and sudden coronary death. Strict control of the dyslipidemia that is frequently found in diabetic patients is essential,

as described below.

Treatment of Dyslipidemia The adverse interactions between the atherogenic lipids (LDL, triglycerides, and lipid remnants) play a critical role in the development of atherosclerosis and the ischemic syndromes. The treatment of dyslipidemia is central when aiming for long-term relief from angina, reduced need for revascularization, and reduction in myocardial infarction and death. Epidemiology, angiographic trials, and controlled trails have shown that (1) men over 45 years and women over 55 years with two risk factors (family history of prematureIHD, cigarette smoking, hypertension, diabetes mellitus) or evidence of atherosclerotic disease should have a total cholesterol£ 5.17 mmol/L (£200 mg/dL), LDL£ 2.58 mmol/L (£100 mg/dL), andHDL³ 1.03 mmol/L (³40 mg/dL); and (2) diabetic patients any age need to achieve the same goals as the likelihood of adverse coronary events is so high. The controlled trials have shown equal benefit for women, the elderly, and even smokers. The control of lipids can be achieved by the combination of a diet low in saturated fatty acids, exercise, and weight loss. Frequently, HMG CoA reductase inhibitors (statins) are required and can lower LDL cholesterol (25 to 60%), raise HDL cholesterol (5 to 9%), and lower triglycerides (5 to 45%). Niacinand fibrates can be used to raise HDL cholesterol and lower triglycerides (Chaps. 242 and341).

Risk reduction in women with IHD The incidence of clinical IHD in premenopausal women is very low. However, following the menopause, the atherogenic risk factors increase (e.g., increasedLDL, reducedHDL) and the rate of clinical coronary events accelerates to the levels observed in men. Women have not given up cigarette smoking as effectively as have men. Diabetes mellitus, which is more common in women, greatly increases the occurrence of clinical IHD and amplifies the deleterious effects of hypertension, hyperlipidemia, and smoking. Cardiac catheterization and coronary revascularization are often applied more sparingly in women and at a later, and more severe, stage of the disease than in men. These factors likely explain the modest increase in complications. Although many of the clinical trials to date have not represented women adequately, the evidence is that when cholesterol lowering, beta blockers after myocardial infarction, andCABG are applied in the appropriate patient groups, women enjoy the same benefits of improved outcome as do men.

Drug Therapy The commonly used drugs for angina pectoris are summarized inTable 244-2.

Nitrates This valuable class of drugs in the management of angina pectoris acts by causing systemic venodilation, thereby reducing myocardial wall tension and oxygen requirements, as well as by dilating the epicardial coronary vessels and increasing blood flow in collateral vessels. The absorption of these agents is most rapid and complete through the mucous membranes. For this reason, nitroglycerin is administered sublingually in tablets of 0.4 or 0.6 mg. Patients with angina should be instructed to take the medication both to relieve angina and also in anticipation of stress (exercise or emotional) that is likely to induce an episode. The value of this prophylactic use of the drug cannot be overemphasized.

Headache and a pulsating feeling in the head are the most common side effects of nitroglycerin and fortunately only rarely become disturbing at the doses usually required

to relieve or prevent angina. Nitroglycerin deteriorates with exposure to air, moisture, and sunlight, so that if the drug neither relieves discomfort or headache nor produces a slight sensation of burning at the sublingual site of absorption, the preparation may be inactive and a fresh supply should be obtained. If relief is not achieved after the first dose of nitroglycerin, a second or third dose may be given at 5-min intervals. If discomfort continues despite treatment, the patient should consult a physician or report promptly to a hospital emergency room for evaluation of possible unstable angina or acute myocardial infarction (Chap. 243).

A diary of angina and nitroglycerin use may be valuable for detecting changes in the frequency or severity of discomfort that may signify the development of unstable angina pectoris and/or herald an impending myocardial infarction.

None of the long-acting nitrates is as effective as sublingual nitroglycerin for the acute relief of angina. These preparations can be swallowed, chewed, or administered as a patch or paste by the transdermal route. They can provide effective plasma levels for up to 24 h, but the therapeutic response is highly variable. Different preparations and/or administration during the daytime should be tried only to prevent discomfort in the individual patient while avoiding side effects such as headache and dizziness. Individual dose titration is important in order to prevent side effects. Useful preparations include isosorbide dinitrate (10 to 60 mg PO bid or tid), nitroglycerin ointment (0.5 to 2.0 in. qid), or sustained-release transdermal patches (5 to 25 mg/d). The nitrates likely bind to guanylate cyclase in vascular smooth muscle cells, oxidize sulfhydryl groups, and are converted to *S*-nitrosothiols. This leads to an increase in cyclic guanosine monophosphate which causes relaxation of vascular smooth muscle. Tolerance with loss of efficacy develops with 12 to 24 h of continuous exposure to all of the long-acting nitrates due to depletion of sulfhydryl groups and to counterregulatory alterations in intravascular fluid balance with fluid retention. In order to minimize the effects of tolerance, the minimum effective dose should be used and a minimum of 8 h each day kept free of the drug so as to restore any useful response(s).

Beta Blockers (See also Chap. 72) These drugs represent an important component of the pharmacologic treatment of angina pectoris. They reduce myocardial oxygen demand by inhibiting the increases in heart rate and myocardial contractility caused by adrenergic activation. Beta blockade reduces these variables most strikingly during exercise while causing only small reductions in heart rate, cardiac output, and arterial pressure at rest. Long-acting beta-blocking drugs (atenolol, 50 to 100 mg/d, and nadolol, 40 to 80 mg/d) offer the advantage of once-a-day dosage (Tables 72-1 and244-2). The therapeutic aims include relief of angina and ischemia. These drugs can also reduce mortality and reinfarction when given to patients after myocardial infarction. Relative contraindications to the use of beta blockers include asthma and reversible airway obstruction in patients with chronic lung disease, atrioventricular conduction disturbances, severe bradycardia, Raynaud's phenomenon, and a history of depression. Side effects include fatigue, impotence, cold extremities, intermittent claudication, bradycardia (sometimes severe), impaired atrioventricular conduction, left ventricular failure, bronchial asthma, and intensification of the hypoglycemia produced by oral hypoglycemic agents and insulin. Reducing the dose or even discontinuation of the drug may be necessary if these side effects develop and persist.

Calcium Antagonists Slow-release nifedipine (30 to 90 mg once daily), verapamil (80 to 120 mg tid), diltiazem (30 to 90 mg qid), amlodipine (2.5 to 10 mg daily), and other calcium antagonists are coronary vasodilators that produce variable and dose-dependent reductions in myocardial oxygen demand, contractility, and arterial pressure. These combined pharmacologic effects are advantageous and make these agents effective in the treatment of angina pectoris. They are indicated when beta blockers are contraindicated, poorly tolerated, or ineffective. Verapamil and diltiazem may produce symptomatic disturbances in cardiac conduction and bradyarrhythmias, exert negative inotropic actions, and are more likely to worsen left ventricular failure, particularly when used in patients with left ventricular dysfunction. Although useful effects are usually achieved when calcium antagonists are combined with beta blockers and nitrates, careful individual titration of dose is essential with these potent combinations. Variant (Prinzmetal's) angina responds particularly well to calcium antagonists, supplemented when necessary by nitrates. Nifedipine as well as other calcium antagonists are now formulated as long-acting preparations including diltiazem (60 to 120 mg twice daily) and verapamil (180 to 240 mg once daily).

Verapamil should not ordinarily be combined with beta blockers because of the combined effects on heart rate and contractility. Diltiazem can be combined with beta blockers with caution and only in patients with normal ventricular function and no conduction disturbances. Nifedipine or amlodipine and the beta blockers have complementary actions on coronary blood supply and myocardial oxygen demands. While the former decreases blood pressure and dilates coronary arteries, the latter slows heart rate and decreases contractility. Nifedipine and the other second-generation dihydropyridine calcium antagonists (nicardipine, isradipine, amlodipine, and felodipine) are potent vasodilators and useful in the simultaneous treatment of angina and hypertension. Short-acting dihydropyridines should be avoided because of the risk of precipitating infarction, particularly in the absence of beta blockers.

Choice between Beta Blockers and Calcium Antagonists for Initial Therapy Since beta blockers have been shown to improve life expectancy following myocardial infarction (p. 1393), they may be preferable in patients with chronicIHD. However, calcium antagonists are indicated in patients with the following: (1) angina and a history of asthma or chronic obstructive pulmonary disease; (2) sick-sinus syndrome or significant atrioventricular conduction disturbances; (3) Prinzmetal's angina; (4) symptomatic peripheral vascular disease; and (5) adverse reactions to beta blockers -- depression, sexual disturbances, fatigue. Many patients with angina do well with a combination of a beta blocker and dihydropyridine calcium antagonist.

Antiplatelet Drugs Aspirin is an irreversible inhibitor of platelet cyclooxygenase activity and thereby interferes with platelet activation. Chronic administration of 100 to 325 mg orally per day has been shown to reduce coronary events in asymptomatic adult men, patients with asymptomatic ischemia after myocardial infarction, patients with chronic stable angina, and patients with or who have survived unstable angina and myocardial infarction. Administration of this drug should be considered in all patients withIHD in the absence of side effects such as gastrointestinal bleeding, allergy, or dyspepsia (Fig. 244-CD6). Clopidogrel is an oral agent that blocks ADP receptor-mediated platelet aggregation. It provides the same benefits as aspirin, if not better, particularly if aspirin causes the side effects listed above.

In summary, a regimen of exercise, smoking cessation, treatment of hypertension and dyslipidemia, aspirin, and beta blockers after infarction are medical interventions that reduce angina, the need for revascularization, myocardial infarction and coronary death.

Treatment af Angina and Heart Failure Transient left ventricular failure with angina can be controlled by the use of nitrates. For patients with established congestive heart failure the increased left ventricular wall tension raises myocardial oxygen demand. Treatment of congestive heart failure with angiotensin-converting enzyme inhibitors, diuretics, and digitalis (Chap. 232) will decrease heart size, wall tension, and myocardial oxygen demands, which, in turn, will help to control angina and ischemia. Nocturnal angina can often be relieved by the treatment of heart failure; however, there is no proven benefit when these drugs are used in patients with angina, a normal heart size, and no evidence of heart failure. Nitrates are particularly useful and can simultaneously improve the disturbed hemodynamics of congestive heart failure by vasodilatation, thereby reducing preload, and relieve angina by preventing or reversing myocardial ischemia. There is some evidence that amlodipine is a calcium antagonist that is well tolerated by patients with left ventricular dysfunction and a valuable agent in the treatment of angina in patients with heart failure. The combination of congestive heart failure and angina in patients withIHDusually indicates a poor prognosis and warrants serious consideration of cardiac catheterization and mechanical revascularization.

CORONARY REVASCULARIZATION

While the basic management of patients withCAD, which is a lifelong condition, is medical, as described above, many patients are improved by coronary revascularization procedures, as described below. These interventions should be employed in conjunction with but do not replace the continuing need to modify risk factors.

PERCUTANEOUS CORONARY INTERVENTION (See also Chap. 245)

PCI, most commonly percutaneous transluminal coronary angioplasty (PTCA) or stenting, is a widely used method to achieve revascularization of the myocardium in patients with symptomaticIHD and suitable stenoses of epicardial coronary arteries (Fig. 244-CD7). Whereas patients with stenosis of the left main coronary artery and those with three-vesselCAD(especially with associated impaired left ventricular function) who require revascularization are best treated withCABG, PCI is widely employed in patients with symptoms and evidence of ischemia due to stenoses of one or two vessels, and even selected patients with three-vessel disease, and may offer many advantages over surgery.

Indications and Patient Selection The most common clinical indication forPCI is angina pectoris, stable or unstable, accompanied by evidence of ischemia in an exercise test. PCI is more effective than medical therapy for the relief of angina. The value of this procedure in reducing the occurrence of coronary death and myocardial infarction has not been established, and therefore it is not generally indicated in asymptomatic or mildly symptomatic patients. PCI can be used to treat stenoses in native coronary arteries as well as in bypass grafts in patients who have recurrent angina following coronary artery surgery. This is an important indication when the

technical difficulties and the increased mortality that accompanies reoperation are considered. PCI has also been carried out in patients with recent total occlusion (within 3 months) of a coronary artery and severe angina; in this group the primary success rate is slightly decreased.

Risks When coronary stenoses are discrete and symmetric, two and three vessels can be dilated in sequence. However, case selection is essential in order to avoid a prohibitive risk of complications. Advanced age, stenoses with thrombus, left ventricular dysfunction, stenosis of an artery perfusing a large segment of myocardium without collaterals, long eccentric or irregular stenoses, and calcified plaques all increase the likelihood of complications but are not absolute contraindications, while left main coronary artery stenosis *is* generally regarded as an absolute contraindication. The major complications are usually due to dissection or thrombosis with vessel occlusion, uncontrolled ischemia, and ventricular failure. Oral aspirin and intravenous heparin are always given to reduce coronary thrombus formation. In unstable angina and when intracoronary thrombus is seen, the use of specific platelet glycoprotein receptor antagonists further reduce thrombotic complications and increase success. In experienced hands, the overall mortality rate should be less than 0.5%, the need for emergency coronary surgery less than 1%, and the occurrence of clinical myocardial infarction less than 2%. Minor complications occur in 5 to 10% of patients and include occlusion of a branch of a coronary artery, myocardial infarction with release of CK-MB into the circulation, and complications of arterial catheterization.

Efficacy Primary success, i.e., adequate dilation (an increase in luminal diameter to a residual diameter obstruction <50%) with relief of angina, is achieved in approximately 95% of cases. Recurrent stenosis of the dilated vessels occurs in 30 to 45% of cases within 6 months ofPTCA, and angina will recur within 6 to 12 months in 25% of cases. This recurrence of symptoms and restenosis is more common in patients with diabetes mellitus, unstable angina, incomplete dilation of the stenosis, dilation of the left anterior descending coronary artery, and stenoses containing thrombi. Dilation of arteries that are totally occluded and of stenotic or occluded vein grafts also exhibits a high incidence of restenosis. It is usual clinical practice to administer aspirin for months after the procedure. Although aspirin and the antiplatelet drug Clopidogrel may help prevent acute coronary thrombosis during and shortly followingPCI, there are no controlled clinical trials that have demonstrated that these medications or any other can clearly reduce the incidence of restenosis. Successful deployment of a metal stent lowers the restenosis rate to 10 to 30% at 6 months but initially requires vigorous antiplatelet therapy (aspirin and Clopidogrel). There is early evidence that local radiation can further reduce restenosis.

If patients do not develop restenosis or angina within the first year after angioplasty, the prognosis for maintaining improvement over the subsequent 4 years is excellent. If restenosis occurs,PTCA can be repeated with the same success and risk, but the likelihood of restenosis increases with the third or subsequent attempt.

SuccessfulPCIproduces effective relief of angina in over 95% of cases and has been shown to be more effective than medical therapy for up to 2 years. Between 30 and 50% of patients with symptomaticIHD who require revascularization can be treated by PCI and need not undergoCABG. Successful PCI is less invasive and expensive than

CABG, usually requires only 1 to 2 days in the hospital, and permits considerable savings in the initial cost of care. Successful PCI also allows earlier return to work and the resumption of an active life. However, this economic benefit is reduced over time because of the greater need for follow-up and for repeat procedures.

CORONARY ARTERY BYPASS GRAFTING

InCABG, a section of a vein (usually the saphenous) is used to form a connection between the aorta and the coronary artery distal to the obstructive lesion. Alternatively, anastomosis of one or both of the internal mammary arteries or a radial artery to the coronary artery distal to the obstructive lesion may be employed and is now preferred whenever possible.

Although some indications for coronary artery bypass surgery are controversial, certain areas of agreement exist:

1. The operation is relatively safe, with mortality rates less than 1% in patients without serious comorbid disease and normal left ventricular function, when the procedure is performed by an experienced surgical team.

2. Intraoperative and postoperative mortality increase with the degree of ventricular dysfunction, comorbidities, age above 80 years, and surgical inexperience. The effectiveness and risk ofCABGvary widely depending on case selection and the skill and experience of the surgical team.

3. Occlusion of vein grafts is observed in 10 to 20% during the first postoperative year and in approximately 2% per year during 5- to 7-year follow-up and 4% per year thereafter. Long-term patency rates are considerably higher for internal mammary and radial artery implantations; in patients with left anterior descending coronary artery obstruction, survival is better when coronary bypass involves the internal mammary artery rather than a saphenous vein. Graft patency and outcomes are improved by meticulous treatment of risk factors, particularly dyslipidemia.

4. Angina is abolished or greatly reduced in approximately 90% of patients following complete revascularization. Although this is usually associated with graft patency and restoration of blood flow, the pain may also have been alleviated as a result of infarction of the ischemic segment or a placebo effect. Within 3 years, angina recurs in about one-fourth of patients but is rarely severe.

5. CABGdoes not appear to reduce the incidence of myocardial infarction in patients with chronicIHD; perioperative myocardial infarction occurs in 5 to 10% of cases, but in most instances these infarcts are small and have little effect on left ventricular function.

6. Mortality is reduced by operation in patients with stenosis of the left main coronary artery as well as in patients with three- or two-vessel disease with significant obstruction of the proximal left anterior descending coronary artery. The survival benefit is greater in patients with abnormal left ventricular function (ejection fraction<50%). Mortality *may* also be reduced in the following patients: (1) with one- or two-vesselCADwithout significant proximal left anterior descending artery CAD but with high-risk criteria on

noninvasive testing; (2) with obstructive CAD who have survived sudden cardiac death or sustained ventricular tachycardia; (3) who have undergone previousCABG and who have multiple saphenous vein graft stenoses, especially of a graft supplying the left anterior descending coronary artery; and (4) with prior PCI recurrent stenosis, and high-risk criteria on noninvasive testing.

Indications forCABG are usually based on the severity of symptoms, coronary anatomy, and ventricular function. The ideal candidate is male, less than 75 years of age, has no other complicating disease, has troublesome or disabling symptoms that are not adequately controlled by medical therapy or does not tolerate medical therapy and wishes to lead a more active life, and has severe stenoses of several epicardial coronary arteries with objective evidence of myocardial ischemia as a cause of the chest discomfort. Great symptomatic benefit can be anticipated in such patients.

Congestive heart failure and/or left ventricular dysfunction (ejection fraction<40%), advanced age (>75 years), reoperation, urgent need for surgery, and the presence of diabetes are all associated with higher perioperative mortality.

Left ventricular dysfunction can be due to noncontractile segments that are viable (hibernating myocardium). These can be detected by using radionuclide scans of myocardial perfusion and metabolism, positron emission tomography, or delayed scanning with thallium-201 or by return of contractile function provoked by low-dose dobutamine. Revascularization can return function and improve survival.

The Choice BetweenPCIandCABG(See Table 244-3) A number of randomized trials have comparedPTCA and CABG in patients with multivesselCAD who were suitable technically for both procedures. The redevelopment of angina requiring repeat coronary angiography and repeat revascularization due to restenosis was higher in the PTCA group. However, the occurrence of death or myocardial infarction has been found to be similar between both groups for up to 5 years. In patients with diabetes plus disease of two or more coronary arteries, bypass surgery results in significantly better outcomes and survival and should be the technique of choice. In addition, the recurrence of angina and stenosis and the need for additional revascularization was much higher in the angioplasty group (about 50%) than in the surgery group (about 10%). Based on these trials and observational studies, we now recommend that patients with an unacceptable level of angina despite optimal medical management should be considered for revascularization. Patients with single- or two-vessel disease with normal or slightly depressed global left ventricular function and anatomically suitable lesions are ordinarily advised initially to undergo PCI (Chap. 245). Patients with two- or three-vessel disease and impaired global left ventricular function (left ventricular ejection fraction <45%) or diabetes mellitus or those with left main disease or other lesions unsuitable for catheter-based procedures should be considered for CABG as the initial method of revascularization (Table 244-3).

UNSTABLE ANGINA PECTORIS

The following three patient groups may be said to have unstable angina pectoris: (1) patients with new onset (<2 months) angina that is severe and/or frequent (³3 episodes per day); (2) patients with accelerating angina, i.e., those with chronic stable angina who

develop angina that is distinctly more frequent, severe, prolonged, or precipitated by less exertion than previously; (3) those with angina at rest. Five mechanisms for unstable angina have been described: (1) a nonocclusive thrombus -- often a platelet plug -- overlying a fissured atherosclerotic plaque; (2) dynamic obstruction -- either spasm of an epicardial coronary artery, as in Prinzmetal's variant angina (see below), or abnormal vasoconstriction of the coronary microcirculation, as in microvascular angina; (3) severe, organic luminal narrowing, as in restenosis following aPCI; (4) arterial inflammation leading to thrombosis; and (5) increase in myocardial oxygen demands caused by conditions such as tachycardia, fever, and thyrotoxicosis in the presence of fixed, severe coronary obstruction. More than one of these may be operative.

When unstable angina is accompanied by objectiveECGevidence of transient myocardial ischemia (ST-segment changes and/or T-wave inversions during episodes of chest pain), it is associated with critical stenoses in one or more major epicardial coronary arteries in about 85% of cases.

TREATMENT

The management of unstable angina is outlined inFig. 244-3. The patient is admitted to the hospital, placed at rest, sedated, and reassured. In all instances, concomitant conditions that can intensify ischemia, such as tachycardia, hypertension, diabetes mellitus, cardiomegaly, heart failure, arrhythmias, thyrotoxicosis, and any acute febrile illness, should be sought and vigorously treated. Acute myocardial infarction should be ruled out by means of serialECGs and measurements of plasma cardiac enzyme activity.

ContinuousECGmonitoring should be carried out. Since thrombus formation frequently complicates this condition, intravenous heparin should be given for 3 to 5 days to maintain the partial thromboplastin time at 2 to 2.5 times control, together with or followed by oral aspirin at a dose of 325 mg/d (Fig. 244-CD8). Alternatively, low-molecular-weight heparin (e.g., enoxaparin, 1 mg/kg subcutaneously b.i.d.) may be used. High-risk unstable angina patients, i.e., those with rest pain, and ST-segment deviations and/or release of a marker of myocardial injury (such as troponin I or T) should also receive an intravenous infusion of a platelet GpIIb/IIIa inhibitor. A beta blocker should be administered and a calcium antagonist added if ischemia persists despite the aforementioned therapy, but with caution and an awareness of the possible side effects discussed above. Dosages of these agents should be raised rapidly, but the patient must be observed carefully to avoid bradycardia, heart failure, and hypotension. Nitroglycerin should be given by the sublingual route as needed for symptoms. Intravenous nitroglycerin is quite effective, especially in patients with episodes of ischemia that are particularly severe or prolonged. It is begun at a dosage of 10 ug/min and is raised in 5 ug/min increments to a level at which chest pain is abolished but systolic arterial pressure is maintained or reduced only slightly and other side effects are avoided. After initial stabilization, either an early invasive strategy (coronary angiography and revascularization) or early conservative strategy (continued medical therapy) can be pursued (Fig. 244-3).

The majority of patients (approximately 80%) improve with rest and medical treatment over a 48-h period. If angina at rest and/orECGevidence of ischemia persist despite 24

to 48 h of the comprehensive treatment described above, then cardiac catheterization and coronary arteriography should be performed in patients with no obvious contraindications for revascularization. If the anatomy is suitable,PCI can be performed. If the coronary anatomy is not suitable for PCI,CABGshould be considered to relieve symptoms and myocardial ischemia and as a means of preventing myocardial damage. The factors that influence the choice between catheter-based and surgical revascularization are similar to those in chronic stable angina.

In the early conservative strategy, if the patient's symptoms and signs are controlled on medical therapy, a diagnostic exerciseECG or perfusion scan or, if exercise is not possible, a pharmacologic stress test (p. 1402) should be carried out near the time of hospital discharge. If there is evidence of severe myocardial ischemia and/or evidence of a high risk of coronary events (p. 1402), consideration should be given to catheterization and, depending on the findings, revascularization. Following discharge, patients with unstable angina should be managed similar to chronic angina patients (p. 1404). Severe obstructiveCAD is often present in patients with unstable angina who respond to medical therapy. Many patients in whom the unstable state is controlled are left with severe chronic stable angina and ultimately require mechanical revascularization.

PRINZMETAL'S VARIANT ANGINA

This relatively uncommon form of unstable angina is characterized by recurrent, prolonged attacks of severe ischemia, caused by episodic focal spasm of an epicardial coronary artery. Approximately three-fourths of patients with Prinzmetal's angina exhibit a mild or moderately severe fixed obstruction (with a luminal diameter 50 to 70% of normal) within 1 cm of the site of spasm. Patients with this condition are often smokers and are younger than patients with unstable angina secondary to coronary atherosclerosis. Ischemic pain usually occurs at rest, sometimes awakens the patient from sleep, and is characterized by multilead ST-segment elevation. The diagnosis may be confirmed by detecting transient spasm occurring spontaneously or following a provocative stimulus (intracoronary acetylcholine, hyperventilation) on coronary arteriography. While long-term survival is excellent, complications include episodes of disabling pain, myocardial infarction, serious ventricular arrhythmias, atrioventricular block, and, rarely, sudden death.

TREATMENT

Management of the acute attack consists of multiple doses of sublingual nitroglycerin, an intravenous infusion of nitroglycerin, and short-acting nifedipine (10 to 30 mg); hypotension should be avoided. In chronic management, long-acting nitrates and calcium antagonists are useful. Beta blockers are of little value, while prazosin, a selective alpha-adrenoceptor blocker, may be useful. Occasionally, mechanical revascularization is helpful in patients with accompanying severe discrete obstructive lesions.

ASYMPTOMATIC (SILENT) ISCHEMIA

ObstructiveCAD, acute myocardial infarction, and transient myocardial ischemia are

frequently asymptomatic. During continuous ambulatoryECGmonitoring, the majority of ambulatory patients with typical chronic stable angina are found to have objective evidence of myocardial ischemia (ST-segment depression) during episodes of chest discomfort while they are active outside the hospital, but many of these patients also appear to have more frequent episodes of asymptomatic ischemia. In addition, there is a large (but as yet unknown) number of totally asymptomatic people with severe coronary atherosclerosis who exhibit ST-segment changes during activity. Some of these patients exhibit higher thresholds to electrically induced pain, others show higher endorphin levels, and still others may be diabetic patients with autonomic dysfunction.

Evidence of frequent episodes of ischemia (symptomatic and asymptomatic) during daily life appears to indicate an increased likelihood of adverse coronary events such as death and myocardial infarction. The widespread use of exerciseECGduring routine examinations has also defined some of these heretofore unrecognized patients with asymptomaticCAD. Longitudinal studies have demonstrated an increased incidence of coronary events (sudden death, myocardial infarction, and angina) in asymptomatic patients with positive exercise tests. In addition, patients with asymptomatic ischemia after suffering a myocardial infarction are at greater risk for a second coronary event.

TREATMENT

The management of patients with asymptomatic ischemia must be individualized. Thus, the physician should consider the following: (1) the degree of positivity of the stress test, particularly the stage of exercise at whichECGsigns of ischemia appear, the magnitude and number of the perfusion defect(s) on thallium scintigraphy, and the change in left ventricular ejection fraction which occurs on radionuclide ventriculography or echocardiography during ischemia and/or during exercise; (2) the ECG leads showing a positive response, with changes in the anterior precordial leads indicating a less favorable prognosis than changes in the inferior leads; and (3) the patient's age, occupation, and general medical condition. Most would agree that an asymptomatic 45-year-old commercial airline pilot with 0.4-mV ST-segment depression in leads V_1 to V_4during mild exercise should undergo coronary arteriography, whereas the asymptomatic, sedentary 75-year-old retiree with 0.1-mV ST-segment depression in leads II and III during maximal activity need not. However, there is no consensus about the appropriate procedure in the large majority of patients for whom the situation is less extreme. Patients with evidence of severe ischemia on noninvasive testing (as outlined earlier) should undergo coronary arteriography. Asymptomatic patients with silent ischemia, three-vesselCAD, and impaired left ventricular function may be considered appropriate candidates forCABG.

The treatment of risk factors, particularly lipid lowering as described above, as well as the use of aspirin and beta blockers have been shown to reduce events and improve outcomes in asymptomatic as well as symptomatic patients with ischemia and provenCAD. While the incidence of asymptomatic ischemia can be reduced by treatment with beta blockers, calcium channel antagonists, and long-acting nitrates, it is not clear whether this is necessary or desirable in patients who have not suffered a myocardial infarction. However, there is evidence that beta-adrenoceptor blockade begun 7 to 35 days after acute myocardial infarction improves survival (Chap. 243).

(Bibliography omitted in Palm version)

Back to Table of Contents

245. PERCUTANEOUS CORONARY REVASCULARIZATION - *Donald S. Baim*

Before 1977, bypass surgery was the only form of revascularization available to treat coronary artery disease. In that year, Andreas Gruntzig performed the first catheter-based coronary revascularization, which he named percutaneous transluminal coronary angioplasty (PTCA). With crude early equipment and limited anatomic capability, fewer than 1000 such procedures were performed worldwide annually until 1981. Through the 1980s and early 1990s, however, progressive improvements in the balloon angioplasty equipment led to improved results, expanded indications for use, and explosive growth in PTCA to the point that in the United States, the annual number of procedures (~300,000) roughly matched the number of surgical bypass operations. This growth has been sustained during the late 1990s with the introduction of a number of newer devices (including stents and atherectomy devices) that further improved the acute success and safety as well as the long-term durability of what is now known more broadly as percutaneous coronary revascularization (PCR) or intervention (PCI). The current annual number of PCRs (~600,000) is thus now greater than the number of coronary bypass operations (~400,000). The dominant role that catheter-based intervention has assumed in the treatment of coronary artery disease has led to definition of the field known as *interventional cardiology*, which now has its own fellowship requirements and Board certification of additional qualifications based on training (a specialized interventional cardiology fellowship beyond basic cardiology training), ongoing experience (75 procedures per year), and a written examination.

All catheter-based coronary interventions are derivatives of diagnostic cardiac catheterization (Chap. 228), in which catheters are introduced into the arterial circulation by needle puncture, advanced into the heart under fluoroscopic guidance, and used for pressure measurements or injections of radio-opaque liquid contrast agents. Interventional procedures differ in that the catheter placed into the ostium of the narrowed coronary artery has a slightly larger diameter, and its lumen is used to convey a flexible, steerable guidewire (diameter <0.5 mm) down the coronary artery lumen, through the narrowing, and into the vessel beyond. This guidewire then serves as the rail over which angioplasty balloons or other therapeutic devices are run to enlarge the narrowed segment of coronary artery (Fig. 245-1). BecausePCR is performed with local anesthesia and requires only a short (1- to 2-day) hospitalization, its use in suitable patients can greatly decrease expense and recovery time compared to those associated with coronary bypass surgery. Not all types of coronary narrowing are well suited to catheter-based intervention, but such intervention is the treatment of choice for roughly 70% of patients with symptomatic single vessel disease and roughly 20% of patients with symptomatic three-vessel disease. Given these anatomic restrictions and the small but definite risk of catheter-based intervention (elective mortality rate 0.4 to 1.0%, compared to a rate of 1 to 3% for elective surgical bypass), PCR should still be viewed as an invasive procedure whose risks and benefits for each individual patient need to be weighed before use. Beyond the decision of which patients should undergo revascularization (versus continued medical management), the selection of which patients should undergo catheter-based rather than surgical revascularization requires detailed understanding of both clinical and coronary angiographic factors, as well as the applicability of various interventional techniques.

INDICATIONS

The main indication forPCRremains the presence of one or more coronary stenoses that are approachable by catheter-based techniques and are thought to be responsible for a clinical syndrome that warrants revascularization. Moreover, the risks and benefits of revascularization by PCR should compare favorably with those of surgery. In patients with significant narrowing of a single coronary artery, the main benefit of revascularization lies in relief of anginal symptoms rather than in increasing their already good prognosis with medical therapy. By contrast, for the patient with significant left main stenosis or multivessel disease, revascularization may both relieve angina *and* improve long-term survival. Most patients with multivessel coronary disease, however, currently undergo surgical rather than catheter-based revascularization, particularly when one or more vessels supplying significant areas of viable myocardium are not well-suited to PCR (owing to chronic total occlusion or other unfavorable anatomic features). In patients for whom either PCR or bypass surgery is a possible treatment for multivessel coronary artery disease, a number of randomized trials have suggested that the two procedures have equivalent in-hospital and 3- to 5-year mortality rates, but that more patients undergoing PCR (40 to 50% versus 7 to 10% surgical patients) will require a second revascularization procedure (generally a repeat PCR to treat restenosis) by 5 years to maintain an equivalent level of symptom relief. One exception may be diabetic patients with multivessel coronary artery disease, for whom some studies have suggested better survival with surgical treatment than with PCR.

The clinical indications forPCRcover the spectrum from patients with unstable angina or acute infarction to patients with silent ischemia, as summarized in the 1999 guidelines. For most patients, PCR is used to treat anatomically approachable lesions that are responsible for the clinical syndrome of *moderately severe, chronic, stable angina*, which persists despite medical antianginal therapy (Chap. 244). Approximately 15% of current patients undergoing PCR, however, have only mild anginal symptoms despite suitable coronary anatomy but have objective evidence of ischemia on noninvasive testing (i.e., an abnormal exercise test). At the other extreme, many patients have more pressing indications for revascularization, including unstable angina or even acute myocardial infarction (with or without prior thrombolytic therapy). An aggressive approach to the treatment of unstable angina involving initial stabilization with beta blockers, nitrates, heparin, and antiplatelet agents (aspirin and frequently a platelet glycoprotein IIb/IIIa receptor blocker), followed by diagnostic catheterization, and same-procedure PCR of the underlying blockage, offers the patient a more rapid return to work, fewer readmissions and late revascularizations, and potentially a reduction in late events compared to prolonged initial trials of medical therapy before proceeding with invasive evaluation and treatment.

In the early 1980s the introduction of intravenous fibrinolytic agents to reopen the occluded infarct-related vessel was a major advance in the treatment of acute myocardial infarction (Chap. 243). It seemed reasonable thatPCRmight further improve the results of thrombolytic therapy by treating the underlying atherosclerotic stenosis, opening those arteries that failed to reperfuse with a thrombolytic alone and preventing the 10 to 20% incidence of in-hospital reocclusion that occurs after even successful thrombolysis. Randomized trials, however, showed that none of the routine PCR strategies tested after thrombolytic administration improved the outcome more than a "watchful waiting" strategy in which PCR was reserved for patients with spontaneous or

exercise-induced ischemia. In contrast, there is evidence that *primary* or direct angioplasty (used instead of thrombolytic therapy) can reduce the in-hospital mortality rate (from roughly 7 to 4%) when performed promptly by a skilled operator (Fig. 245-2). Another advantage of PCR is that it can be performed even in the approximately 30% of patients with acute myocardial infarction with contraindications to thrombolytic therapy.

As the clinical indications forPCR have broadened, so have its anatomic capabilities. PCR thus is no longer restricted to proximal, discrete, subtotal, concentric, noncalcified lesions, as was the case initially. Calcified, complex, or diffuse disease lesions respond well to coronary stent placement, sometimes after pretreatment with rotational atherectomy. Even totally occluded coronary arteries (particularly ones that have been occluded for less than 6 months) can be crossed and dilated effectively, although the success rate remains somewhat lower than for subtotal lesions (60% versus 90% for subtotal stenotic lesions). In addition to lesions in the native coronary tree, obstructions in saphenous vein (Fig. 245-3) or internal mammary artery bypass grafts also can be dilated successfully to treat postbypass angina. If multiple lesions are responsible for the clinical syndrome, they generally can be dilated during a single procedure.

RESULTS

The success rate forPCRexceeds 95% for dilating a target lesion so that its residual diameter stenosis is <50% (<30% when a stent has been used), without producing an associated complication. About half of the failures result from inability to cross the target lesion with the guidewire or balloon catheter, particularly when that target lesion is a chronic total occlusion. With balloon dilatation alone, some local dissection is present in virtually all successful procedures. Before the introduction of stent technology (see below), more extensive dissection (particularly in association with local thrombus formation or vasospasm) led to abrupt closure of the dilated segment soon after withdrawal of the balloon catheter, and necessitated emergency bypass surgery in approximately 3% of angioplasty attempts. However, such dissections are now routinely treated by stent placement, reducing the incidence of emergency bypass surgery to <1%. Other than dissection, the main hazards of PCR concern spasm, thrombosis, and perforation.

Coronary spasm is controlled by the routine use of vasodilators (nitrates and calcium channel antagonists), whereas thrombosis is controlled by systemic anticoagulation (heparin, 7000 to 10,000 units during the procedure to maintain an activated clotting time of 250 to 300 s), and antiplatelet therapy (aspirin, 325 mg/d starting at least 24 h beforePCR and continued for at least 3 to 6 months after the procedure). If a coronary stent has been placed, aspirin is supplemented by a blocker of the platelet ADP receptor (ticlopidine or clopidogrel) to reduce the likelihood of stent thrombosis (see below). Newer potent intravenous antiplatelet agents (blockers of the platelet glycoprotein IIb/IIIa receptors) may reduce further the incidence of ischemic complications within 72 h of PCR, and are used prophylactically in what are perceived to be high-risk interventions or provisionally in interventions that have left behind an imperfect mechanical result (e.g., an unstented distal dissection).

Perforation of a coronary artery was an extremely rare complication of conventional balloon angioplasty but may occur in up to 1% of patients undergoing more aggressive

atherectomy procedures (see below). Even small perforations of the distal vessel by the angioplasty guidewire may lead to significant hemopericardium requiring urgent pericardiocentesis in the setting of intense anticoagulant and antiplatelet therapy. Finally, catheter-based interventions are subject to all of the complications of diagnostic catheterization, including adverse reactions to iodinated contrast agents and groin hematoma. By and large, however, catheter-based coronary revascularization has reached the point of being a safe and effective alternative to surgical revascularization.

FOLLOW-UP

After successfulPCR of all "culprit" lesions, marked improvement or complete resolution of the presenting ischemic syndrome should be evident. In approximately 20% of patients, however, evidence of recurrent ischemia develops within 6 months, due to restenosis of the dilated segment. This restenosis appears to result from excessive local fibrointimal proliferation and vessel constriction, occurring in response to the local injury that is part of enlarging the stenotic lumen. When recurrent ischemia develops more than 6 months after PCR, it usually reflects progression of disease at another site, rather than restenosis. Whether due to restenosis or disease progression, most post-PCR problems can be treated by repeat PCR, so that only about 10% of patients require bypass surgery during the 5 years after a successful procedure. When a patient has provided evidence of severe obstructive coronary atherosclerosis requiring revascularization, either by bypass surgery or PCR, the opportunity to implement an aggressive program to reduce atherosclerotic risk factors and thereby slow the pace of development of new lesions should not be overlooked (Chap. 244).

NONBALLOON TECHNIQUES

Conventional balloon angioplasty(PTCA)was the only catheter-based coronary revascularization technique that was widely available before 1990. Although it offered anatomic versatility and acceptable short- and long-term results, the difficulty of using this technique for certain anatomic lesion types (e.g., calcified eccentric, ostial, thrombus-containing, or bifurcation lesions) and the persistence of problems such as abrupt closure and restenosis fostered the development of a number of newer, nonballoon techniques that include stent placement and atherectomy. These treatments moved from clinical investigation to routine clinical practice during the early 1990s and now account for 70 to 80% of percutaneous coronary interventions. Used appropriately, these new techniques have improved the success, safety, and long-term results (restenosis rate) in most lesion types. Most of these procedures cost more than PTCA, but much of this cost can be recouped by the reduction in long-term expenses for the treatment of restenosis. Given these developments, stand-alone balloon angioplasty is now used in a minority of procedures (20% of allPCRs), although adjunctive balloon angioplasty is still routinely used to pre- or postdilate, before or after a newer interventional device.

STENTS

Stents are metallic scaffolds that are inserted into a diseased vessel segment in their collapsed form and are then expanded (by balloon expansion, or by self-expansion after removal of a constraining membrane) to establish a normal-appearing vessel lumen

(Fig. 245-CD1). Stents overcome two of the principal limitations of balloon dilatation -- the tendency for elastic recoil of the vessel wall and local dissection of the plaque. As such, stents provide a larger acute lumen than does conventional balloon angioplasty, which allows them to reduce the incidence of subsequent restenosis by roughly one-third (e.g., angiographic restenosis rates of 20% versus 33%, and clinical restenosis rates of 10% versus 16 to 20%). When in-stent restenosis does occur, it is almost never the result of stent crush but rather the consequence of excessive neointimal hyperplasia within the stent (Fig. 245-4). In-stent restenosis can be treated by atherectomy to remove the excess tissue (see below), balloon dilatation, and then local delivery of b org radiation to suppress neointimal regrowth.

Two balloon-expandable stent designs were approved by the Food and Drug Administration (FDA) in the early 1990s -- a wire coil design for use in stabilizing actual or threatened abrupt closure and a slotted tube design for elective treatment of native coronary lesions. After their release, the efficacy of the slotted tube design was demonstrated in a variety of other circumstances, including restenotic lesions and saphenous vein grafts (Fig. 245-3). In the late 1990s, a number of second generation stent designs were developed that offer easier delivery to tortuous or distal lesions as well as a wider variety of sizes and lengths. The approval of these devices has allowed them to completely replace the first generation devices in clinical practice (Fig. 245-5). Still further refinements in stent coverings (to seal aneurysms or perforations) and coatings (to suppress stent thrombosis and in-stent proliferation) are in progress.

Early experience suggested that metallic stents were prone to thrombotic occlusion, either acute (<24 h) or subacute (1 to 14 days with a peak at 6 days), and that an aggressive anticoagulation regimen (aspirin, dipyridamole, and warfarin) was needed to prevent such thrombosis. This aggressive anticoagulant regimen reduced the incidence of stent thrombosis to ~3% but led to longer hospitalization and an increased incidence of local vascular complications at the femoral arterial entry site. Subsequent data suggested that many of these thrombotic complications were the result of incomplete stent expansion and that more attention to full initial deployment would allow the same stents to be used with only antiplatelet drugs (aspirin plus the platelet ADP-receptor blockers, ticlopidine or clopidogrel) with more acceptable thrombosis and vascular complication rates (each<1%). This rapid evolution in devices, concomitant medications, and indications has led to the dominance of stent placement in catheter-based coronary revascularization, with placement of one or more stents in 70 to 80% of all procedures.

Atherectomy Whereas both balloon angioplasty and stent placement enlarge the coronary lumen by displacing plaque, atherectomy catheters enlarge the lumen by removing plaque mass from the treated lesion. Directional atherectomy achieves this result by use of a special catheter with a windowed steel cylinder at its tip. Inflation of a low-pressure positioning balloon on the back of the cylinder presses plaque into the window, where it is cut and trapped by a spinning cup-shaped cutter. This device was the first (1990) approved nonballoon technology to reach clinical practice, and it is still the treatment of choice for noncalcified lesions at the origin of the left anterior descending artery or at major coronary bifurcations (Fig. 245-6). Although its efficacy over conventional balloon angioplasty has been demonstrated, the ease and result of stent placement are much greater for most other lesion types. Rotational atherectomy uses burrs of various sizes (diameter 1.25 to 2.50 mm) that are coated on their leading

half with small diamond chips. The burr is spun at 140,000 to 160,000 rpm as it is advanced through a coronary lesion over a leading guidewire. As the burr is advanced, the diamond chips grind through the obstructing plaque, and pulverize it into small (5 to 25 um) particles, which pass through the distal coronary microcirculation. This device has emerged as an effective treatment for long (>20 mm), calcified, ostial lesions or in-stent restenotic lesions, frequently followed by balloon dilation or stent placement. *Extraction* atherectomy uses a combination of distal cutting blades rotating at low speed and continuous vacuum aspiration to remove coronary obstructions. The device has limited cutting efficiency, and its use is now confined to softer lesions (e.g., atherosclerotic saphenous vein grafts) or thrombotic lesions. Newer aspiration devices based on the Bernoulli principle appear better able to remove clot and cause less vessel disruption.

Although it is not mechanical, laser light [at wavelengths from the ultraviolet (308 mm) to the midinfrared (2000 um)] can be delivered to obstructing coronary plaques through bundles of small optical fibers housed in flexible catheters whose outer diameter is between 1.2 and 2.0 mm. When these catheters are pulsed with laser energy as they are advanced through a coronary obstruction over a guidewire, they can ablate noncalcified coronary plaque by a combination of photoacoustic (blast), thermal, and photochemical effects. Although lasers have been used to treat ostial as well as diffuse coronary lesions, acceptance of the technique has been limited by the expense of the device and the fact that these lesions can be treated by other techniques, such as rotational atherectomy.

SUMMARY

With the development of new techniques such as stent placement and atherectomy, new drug regimens, and a preponderance of "evidence-based" practices, over the last 20 years catheter-based revascularization (PCR) has developed from a procedural curiosity to one of the mainstays of coronary revascularization. As short- and long-term results have improved and the number of procedures has continued to grow, the pace of development, if anything, has intensified.

(Bibliography omitted in Palm version)

Back to Table of Contents

246. HYPERTENSIVE VASCULAR DISEASE - *Gordon H. Williams*

An elevated arterial pressure is probably the most important public health problem in developed countries. It is common, asymptomatic, readily detectable, usually easily treatable, and often leads to lethal complications if left untreated (Chap. 35). As a result of extensive educational programs in the late 1960s and 1970s by both private and government agencies, the number of undiagnosed and/or untreated patients was reduced significantly by the late 1980s to a level of about 25%, with a concomitant decline in cardiovascular mortality. Unfortunately, by the mid-1990s, this beneficial trend began to wane. The number of undiagnosed patients with hypertension increased to nearly 33%, the decline in cardiovascular mortality flattened, and the number of individuals with chronic diseases with untreated or poorly treated hypertension increased. For example, the prevalence of end-stage renal disease per million population increased from <100 in 1982 to>250 in 1995, and the prevalence of congestive heart failure from ages 55 to 75 more than doubled between 1976 to 1980 and 1988 to 1991. Thus, although our understanding of the pathophysiology of elevated arterial pressure has increased, in 90 to 95% of cases the etiology (and thus potentially the means of prevention or cure) is still largely unknown. As a consequence, in most cases the hypertension is treated nonspecifically, resulting in a large number of minor side effects and a relatively high (50 to 60%) noncompliance rate.

PREVALENCE

The prevalence of hypertension depends on both the racial composition of the population studied and the criteria used to define the condition. In a white suburban population like that in the Framingham Study, nearly one-fifth of individuals have blood pressures >160/95, while almost one-half have pressures >140/90. An even higher prevalence has been documented in the nonwhite population. In females the prevalence is closely related to age, with a substantial increase occurring after age 50. This increase is presumably related to the hormonal changes of menopause, although the mechanism is unclear. Thus, the ratio of hypertension frequency in women versus men increases from 0.6 to 0.7 at age 30 to 1.1 to 1.2 at age 65.

The prevalence of various forms of secondary hypertension depends on the nature of the population studied and on how extensive the evaluation is. There are no available data to define the frequency of secondary hypertension in the general population, although in middle-aged males it has been reported to be 6%. On the other hand, in referral centers where patients undergo an extensive evaluation, it has been reported to be as high as 35%. The various forms of hypertension are outlined in Table 246-1, and their relative frequencies are given in Table 246-2.

ESSENTIAL HYPERTENSION

Patients with arterial hypertension and no definable cause are said to have *primary*, *essential*, or *idiopathic hypertension*. Undoubtedly, the primary difficulty in uncovering the mechanism(s) responsible for the hypertension in these patients is attributable to the variety of systems that are involved in the regulation of arterial pressure -- peripheral and/or central adrenergic, renal, hormonal, and vascular -- and to the complexity of the interrelations of these systems. Several abnormalities have been described in patients

with essential hypertension, often with a claim that one or more of them are primarily responsible for the hypertension. While it is still uncertain whether these individual abnormalities are primary or secondary, varying expressions of a single disease process or reflective of separate disease entities, the accumulating data increasingly support the latter hypothesis. Therefore, just as pneumonia is caused by a variety of infectious agents, even though the clinical picture observed may be similar, so essential hypertension likely has a number of distinct causes. Thus, the distinction between primary and secondary hypertension has become blurred, and the approach to both the diagnosis and therapy of hypertensive patients has been modified. For example, when a group of patients with essential hypertension is separated into a distinct subset (e.g., low-renin essential hypertension), the patients have not been reclassified as having a form of secondary hypertension but rather remain in the essential hypertensive group. In this**chapter**, individuals in whom a specific structural organ or gene defect is responsible for hypertension are defined as having a *secondary* form of hypertension. In contrast, individuals in whom generalized or functional abnormalities may be the cause of hypertension, even if the abnormalities are discrete, are defined as having *essential* hypertension.

GENETIC CONSIDERATIONS

Genetic factors have long been assumed to be important in the genesis of hypertension. Data supporting this view can be found in animal studies as well as in population studies in humans. One approach has been to assess the correlation of blood pressure in families (familial aggregation). From these studies, the minimum size of the genetic factor can be expressed by a correlation coefficient of approximately 0.2. However, the variation in the size of the genetic factor in different studies reemphasizes the probably heterogeneous nature of the essential hypertensive population. In addition, most studies support the concept that the inheritance is probably multifactorial or that a number of different genetic defects each have an elevated blood pressure as one of their phenotypic expressions. Finally, both monogenic defects (e.g., glucocorticoid-remediable aldosteronism and Liddle's syndrome) and susceptibility genes (e.g., the angiotensinogen and a adducin genes) have now been reported which have as one of their consequences an increased arterial pressure (see below and Chap. 331). Yet, as can be seen inTable 246-3, most studies of likely genes have failed to document linkage or consistent association with hypertension. However, uncertainty exists as to the validity of these negative conclusions. A positive relationship between hypertension and a gene could be obscured by the high probability of a false-negative result because of the heterogeneity of the hypertensive population. Thus, intermediate phenotypes in the hypertensive population need to be identified to differentiate patients into more homogeneous subgroups; the role of a specific candidate gene can then be more readily assessed. Such an approach is illustrated inTable 246-4.

ENVIRONMENT

A number of environmental factors have been implicated in the development of hypertension, including salt intake, obesity, occupation, alcohol intake, family size, and crowding. These factors have all been assumed to be important in the increase in blood pressure with age in more affluent societies, in contrast to the decline in blood pressure with age in less affluent groups.

SALT SENSITIVITY

The environmental factor that has received the greatest attention is salt intake. Even this factor illustrates the heterogeneous nature of the essential hypertensive population, in that the blood pressure in only approximately 60% of hypertensives is particularly responsive to the level of sodium intake. The cause of this special sensitivity to salt varies, with primary aldosteronism, bilateral renal artery stenosis, renal parenchymal disease, and low-renin essential hypertension accounting for about half the patients. In the remainder, the pathophysiology is still uncertain, but postulated contributing factors include chloride intake, calcium intake, a generalized cellular membrane defect, insulin resistance, and "nonmodulation" (see below).

ROLE OF RENIN

Renin is an enzyme secreted by the juxtaglomerular cell of the kidney and linked with aldosterone in a negative feedback loop (Chap. 331). While a variety of factors can modify its rate of secretion, the primary determinant is the volume status of the individual, particularly as related to changes in dietary sodium intake. The end product of the action of renin on its substrate is the generation of the peptide angiotensin II. The response of target tissues to this peptide is uniquely determined by the prior dietary electrolyte intake. For example, sodium intake normally modulates adrenal and renal vascular responses to angiotensin II. With sodium restriction, adrenal responses are enhanced and the renal vascular responses reduced. Sodium loading has the opposite effect. The range of plasma renin activities observed in hypertensive subjects is broader than in normotensive individuals. In consequence, some hypertensive patients have been defined as having *low-renin* and others as having *high-renin* essential hypertension.

Low-Renin Essential Hypertension Approximately 20% of patients who by all other criteria have essential hypertension have suppressed plasma renin activity. This situation is more common in individuals of African descent than in white patients. Though these patients are not hypokalemic, they have been reported to have expanded extracellular fluid volumes, and it has been suggested but not proved that they have sodium retention and renin suppression due to excessive production of an unidentified mineralocorticoid. On the other hand, some, but not all, studies have suggested that the adrenal cortex of some of these patients has an increased sensitivity to angiotensin II as the underlying mechanism. Not only does this hypothesis potentially explain their low plasma renin activity, it also suggests the cause of their hypertension. On a diet with a normal or high sodium content, aldosterone production will not be suppressed normally, leading to a mild degree of hyperaldosteronism with its resulting increased sodium retention, volume expansion, and increase in blood pressure. Since this altered sensitivity has been reported even in patients with normal-renin hypertension, it is likely that patients with low-renin hypertension are not a distinct subset but rather form part of a continuum of patients with essential hypertension.

Nonmodulating Essential Hypertension Another subset of hypertensive patients has an adrenal defect opposite to that observed in some low-renin patients -- a reduced adrenal response to sodium restriction. In these individuals, sodium intake does not

modulate either adrenal or renal vascular responses to angiotensin II. Hypertensives in this subset have been termed *nonmodulators* because of the absence of the sodium-mediated modulation of target tissue responses to angiotensin II. These individuals make up 25 to 30% of the hypertensive population, have plasma renin activity levels that are normal to high if measured when the patient is on a low-salt diet, and have hypertension that is salt-sensitive because of a defect in the kidney's ability to excrete sodium appropriately. They also are more insulin-resistant than ohter hypertensive patients, and the pathophysiologic characteristics can be corrected by the administration of a converting-enzyme inhibitor. Furthermore, the nonmodulation characteristic appears to be genetically determined (associated with a certain allele of the angiotensinogen gene). Thus, nonmodulators are probably the most completely characterized intermediate phenotype in the hypertensive population.

High-Renin Essential Hypertension Approximately 15% of patients with essential hypertension have plasma renin activity levels above the normal range. It has been suggested that plasma renin plays an important role in the pathogenesis of the elevated arterial pressure in these patients. However, most studies have found that saralasin (a substance that, like losartan, acts as a competitive antagonist of angiotensin II) significantly reduces blood pressure in fewer than half of these patients. This finding has led some investigators to postulate that the elevated renin levels and blood pressure may both be secondary to an increase in adrenergic system activity. It has been proposed that, in patients with angiotensin-dependent high-renin hypertension whose arterial pressures are lowered by an angiotensin II antagonist, the mechanism responsible for the increase in renin and, therefore, for the hypertension is the nonmodulating defect.

SODIUM ION VERSUS CHLORIDE OR CALCIUM

Most studies assessing the role of salt in the hypertensive process have assumed that it is the sodium ion that is important. However, some investigators have suggested that the chloride ion may be equally important. This suggestion is based on the observation that feeding chloride-free sodium salts to salt-sensitive hypertensive animals fails to increase arterial pressure. Calcium has also been implicated in the pathogenesis of some forms of essential hypertension. A low-calcium intake has been associated with an increase in blood pressure in epidemiologic studies; an increase in leukocyte cytosolic calcium levels has been reported in some hypertensives. Finally, calcium entry blockers are effective antihypertensive agents. Several studies have reported a potential link between the salt-sensitive forms of hypertension and calcium. It has been postulated that salt loading in combination with a defect in the kidney's ability to excrete salt may lead to a secondary increase in circulating natriuretic factors. One of these factors, the so-called digitalis-like natriuretic factor, inhibits ouabain-sensitive Na^+, K^+-ATPase and thereby leads to intracellular calcium accumulation and a hyperreactive vascular smooth muscle.

CELL MEMBRANE DEFECT

Another postulated explanation for salt-sensitive hypertension is a generalized cell membrane defect. This hypothesis derives most of its data from studies on circulating blood elements, particularly red blood cells, in which abnormalities in the transport of

sodium across the cell membrane have been documented. Since both increases and decreases in the activity of different transport systems have been reported, it is likely that some abnormalities are primary and some are secondary. It has been assumed that this abnormality in sodium transport reflects an undefined alteration in the cell membrane and that this defect occurs in many, perhaps all, cells of the body, particularly the vascular smooth-muscle cells. The defect leads to an abnormal accumulation of calcium in vascular smooth muscle, resulting in a heightened vascular responsiveness to vasoconstrictor agents. This defect has been proposed to be present in 35 to 50% of essential hypertensive persons on the basis of studies using red cells. Other studies suggest that the abnormality in red cell sodium transport is not fixed but can be modified by environmental factors.

The common final pathway in all these hypotheses is an increase in cytosolic calcium resulting in increased vascular reactivity. However, as described above, several mechanisms might produce this calcium accumulation.

INSULIN RESISTANCE

Insulin resistance and/or hyperinsulinemia have been suggested as being responsible for the increased arterial pressure in some patients with hypertension. While it is clear that a substantial fraction of the hypertensive population has insulin resistance and hyperinsulinemia, it is less certain that this is more than an association. Insulin resistance is common in patients with non-insulin-dependent diabetes mellitus (NIDDM) or obesity. Both obesity and NIDDM are more common in hypertensive than in normotensive subjects. However, several studies have found that hyperinsulinemia and insulin resistance are present even in lean hypertensive patients without NIDDM, suggesting that this relationship is more than a coincidence. As noted earlier, these individuals seem to be concentrated in the nonmodulation phenotype.

Hyperinsulinemia can increase arterial pressure by one or more of four mechanisms. An underlying assumption in each case is that some, but not all, of the target tissues of insulin are resistant to its effects. Specifically, tissues involved in glucose homeostasis are resistant (thereby producing the hyperinsulinemia), while tissues involved in the hypertensive process are not. First, hyperinsulinemia produces renal sodium retention (at least acutely) and increases sympathetic activity. Either or both of these effects could lead to an increase in arterial pressure. Another mechanism is vascular smooth-muscle hypertrophy secondary to the mitogenic action of insulin. Third, insulin also modifies ion transport across the cell membrane, thereby potentially increasing the cytosolic calcium levels of insulin-sensitive vascular or renal tissues. This mechanism would increase arterial pressure for reasons similar to those described above for the membrane-defect hypothesis. Finally, insulin resistance may be a marker for another pathologic process, e.g., nonmodulation, which could be the primary mechanism increasing blood pressure. It is important to point out, however, that the role of insulin in controlling arterial pressure is only vaguely understood, and, therefore, its potential as a pathogenic factor in hypertension remains unclear.

Few of the features of hypertension discussed above remain constant in a given patient. Some may be a reflection of the current metabolic and hormonal status of the patient rather than a permanent feature of the disease process. For example, at one point a

patient might have insulin resistance secondary to obesity, which could lead to sodium retention, intravascular volume expansion, and renin suppression. This patient would be labeled as having "low-renin essential hypertension." If the patient lost weight, however, the salt-retaining tendency would be reversed. If the blood pressure did not normalize, the patient might then have "normal or high-renin essential hypertension." Thus, the features reviewed above should not be considered mutually exclusive or permanent characteristics in a given patient with hypertension.

FACTORS THAT MODIFY THE COURSE OF ESSENTIAL HYPERTENSION

Age, race, sex, smoking, alcohol intake, serum cholesterol, glucose intolerance, and weight may all alter the prognosis of this disease. The younger the patient when hypertension is first noted, the greater is the reduction in life expectancy if the hypertension is left untreated. In the United States, urban blacks have about twice the prevalence of hypertension as whites and more than four times the hypertension-induced morbidity rate. At all ages and in both white and nonwhite populations, females with hypertension fare better than males up to the age of 65, and the prevalence of hypertension in premenopausal females is substantially less than that in age-matched males or postmenopausal women. Yet, compared with their normotensive counterparts, females with hypertension run the same relative risk of a morbid cardiovascular event as males do. Accelerated atherosclerosis is an invariable companion of hypertension. Thus, it is not surprising that independent risk factors associated with the development of atherosclerosis, such as an elevated serum cholesterol, glucose intolerance, and/or cigarette smoking, significantly enhance the effect of hypertension on mortality rate regardless of age, sex, or race (Chap. 241). There also is no question that a positive correlation exists between obesity and arterial pressure. A gain in weight is associated with an increased frequency of hypertension in persons with normal blood pressure, and weight loss in obese persons with hypertension lowers their arterial pressure and, if they are being treated for hypertension, the intensity of therapy required to keep them normotensive. Whether these changes are mediated by changes in insulin resistance is unknown.

NATURAL HISTORY

Because essential hypertension is a heterogeneous disorder, variables other than the arterial pressure modify its course. Thus, the probability of developing a morbid cardiovascular event with a given arterial pressure may vary as much as 20-fold depending on whether associated risk factors are present (Table 246-5). Although exceptions have been reported, most untreated adults with hypertension will develop further increases in arterial pressure with time. Furthermore, it has been demonstrated from both actuarial data and experience in the era prior to effective therapy that untreated hypertension is associated with a shortening of life by 10 to 20 years, usually related to an acceleration of the atherosclerotic process, with the rate of acceleration in part related to the severity of the hypertension. Even individuals who have relatively mild disease -- i.e., without evidence of end organ damage -- that is left untreated for 7 to 10 years have a high risk of developing significant complications. Nearly 30% will exhibit atherosclerotic complications, and more than 50% will have end organ damage related to the hypertension itself, such as cardiomegaly, congestive heart failure, retinopathy, a cerebrovascular accident, and/or renal insufficiency. Thus, even in its mild forms,

hypertension is a progressive and lethal disease if left untreated.

SECONDARY HYPERTENSION

As noted earlier, in only a small minority of patients with elevated arterial pressure can a specific cause be identified. Yet these patients should not be ignored for at least two reasons: (1) correction of the cause may cure their hypertension, and (2) these secondary forms of the disease may provide insight into the etiology of essential hypertension. Nearly all the secondary forms of hypertension are related to an alteration in hormone secretion and/or renal function and are discussed in detail in other chapters.

RENAL HYPERTENSION (See also Chap. 278)

Hypertension produced by renal disease is the result of either (1) a derangement in the renal handling of sodium and fluids leading to volume expansion or (2) an alteration in renal secretion of vasoactive materials resulting in a systemic or local change in arteriolar tone. The main subdivisions of renal hypertension are renovascular hypertension, including preeclampsia and eclampsia, and renal parenchymal hypertension. A simple explanation for *renal vascular hypertension* is that decreased perfusion of renal tissue due to stenosis of a main or branch renal artery activates the renin-angiotensin system, described in Chap. 331. Circulating angiotensin II elevates arterial pressure by directly causing vasoconstriction, by stimulating aldosterone secretion with resulting sodium retention, and/or by stimulating the adrenergic nervous system. In practice, only about one-half of patients with renovascular hypertension have an absolute elevation in renin activity in peripheral plasma, although when renin measurements are referenced against an index of sodium balance, a much higher fraction have inappropriately high values.

Activation of the renin-angiotensin system has also been offered as an explanation for the hypertension in both acute and chronic *renal parenchymal disease*. In this formulation, the only difference between renovascular and renal parenchymal hypertension is that the decreased perfusion of renal tissue in the latter case results from inflammatory and fibrotic changes involving multiple small intrarenal vessels. There are enough differences between the two conditions, however, to suggest that other mechanisms are active in renal parenchymal disease. Specifically, (1) peripheral plasma renin activity is elevated far less frequently in renal parenchymal than in renovascular hypertension; (2) cardiac output is said to be normal in renal parenchymal hypertension (unless uremia and anemia are present) but slightly elevated in renovascular hypertension; (3) circulatory responses to tilting and to the Valsalva maneuver are exaggerated in the latter condition; and (4) blood volume tends to be high in patients with severe renal parenchymal disease and low in patients with severe unilateral renovascular hypertension. Alternative explanations for the hypertension in renal parenchymal disease include the possibilities that the damaged kidneys (1) produce an unidentified vasopressor substance other than renin, (2) fail to produce a necessary humoral vasodilator substance (perhaps prostaglandin or bradykinin), (3) fail to inactivate circulating vasopressor substances, and/or (4) are ineffective in disposing of sodium. In the last case, the retained sodium would be responsible for the hypertension as outlined earlier. Although all these explanations, including participation of the renin-angiotensin system, probably have some validity in individual patients, the

hypothesis involving sodium retention is particularly attractive. It is supported by the observation that those patients with chronic pyelonephritis or polycystic renal disease who are salt wasters do not develop hypertension and by the observation that removal of salt and water by dialysis or diuretics is effective in controlling arterial pressure in most patients with renal parenchymal disease.

A rare form of renal hypertension results from the excess secretion of renin by juxtaglomerular cell tumors or nephroblastomas. The initial presentation is similar to that of hyperaldosteronism, with hypertension, hypokalemia, and overproduction of aldosterone. However, in contrast to primary aldosteronism, peripheral renin activity is *elevated instead of subnormal*. This disease can be distinguished from other forms of secondary aldosteronism by the presence of normal renal function and unilateral increases in renal vein renin concentration without a renal artery lesion.

ENDOCRINE HYPERTENSION

Adrenal Hypertension Hypertension is a feature of a variety of adrenal cortical abnormalities. In *primary aldosteronism* (Chap. 331), there is a clear relationship between the aldosterone-induced sodium retention and the hypertension. Normal individuals given aldosterone develop hypertension only if they also ingest sodium. Since aldosterone causes sodium retention by stimulating renal tubular exchange of sodium for potassium, hypokalemia is a prominent feature in most patients with primary aldosteronism, and, therefore, the measurement of serum potassium provides a simple screening test. The effect of sodium retention and volume expansion in chronically suppressing plasma renin activity is critical for the definitive diagnosis. In most clinical situations, plasma renin activity and plasma or urinary aldosterone levels parallel each other, but in patients with primary aldosteronism, aldosterone levels are high and relatively fixed because of autonomous aldosterone secretion, whereas plasma renin activity levels are suppressed and respond sluggishly to sodium depletion. Primary aldosteronism may be secondary to either a tumor or bilateral adrenal hyperplasia. It is important to distinguish between these two conditions preoperatively, since the hypertension in the latter case is usually not modified by operation.

The sodium-retaining effect of large amounts of glucocorticoids (perhaps resulting in part from saturation of the 11b-hydroxysteroid hydrogenase enzyme system in the kidney by the increased concentration of cortisol) also offers an explanation for the hypertension in severe cases of Cushing's syndrome (Chap. 331). Moreover, increased production of mineralocorticoids has also been documented in some patients with Cushing's syndrome. However, the hypertension in many cases of Cushing's syndrome does not seem volume-dependent, leading investigators to speculate that it may be secondary to glucocorticoid-induced production of renin substrate (angiotensin-mediated hypertension). In the forms of the adrenogenital syndrome due to C-11 or C-17 hydroxylase deficiency (Chap. 331), deoxycorticosterone accounts for the sodium retention and the resulting hypertension, which is accompanied by suppression of plasma renin activity.

In patients with pheochromocytoma (Chap. 332), increased secretion of epinephrine and norepinephrine by a tumor (most often located in the adrenal medulla) causes excessive stimulation of adrenergic receptors, which results in peripheral vasoconstriction and

cardiac stimulation. This diagnosis is confirmed by demonstrating increased urinary excretion of epinephrine and norepinephrine and/or their metabolites.

Acromegaly (See also Chap. 328) Hypertension, coronary atherosclerosis, and cardiac hypertrophy are frequent complications of this condition.

Hypercalcemia (See alsoChap. 340) The hypertension that occurs in up to one-third of patients with hyperparathyroidism ordinarily can be attributed to renal parenchymal damage due to nephrolithiasis and nephrocalcinosis. However, increased calcium levels can also have a direct vasoconstrictive effect. In some cases, the hypertension disappears when the hypercalcemia is corrected. Thus, paradoxically, the increased serum calcium level in hyperparathyroidism raises blood pressure, while epidemiologic studies suggest that a high calcium intake lowers blood pressure. To further confuse the issue, calcium entry-blocking agents are effective antihypertensive agents. Additional studies are needed to resolve these seemingly conflicting observations.

Oral Contraceptives Several years ago, a common cause of endocrine hypertension was the use of estrogen-containing oral contraceptives. However, several studies have since suggested that this is no longer true, probably owing to the lower estrogen content of modern oral contraceptives. In patients receiving these agents who do become hypertensive, the mechanism is likely to be activation of the renin-angiotensin-aldosterone system. Thus, both volume (aldosterone) and vasoconstrictor (angiotensin II) factors are important. The estrogen component of oral contraceptive agents stimulates the hepatic synthesis of the renin substrate angiotensinogen, which in turn favors the increased production of angiotensin II and secondary aldosteronism. Some women taking oral contraceptives have increased plasma concentrations of angiotensin II and aldosterone with some increase in arterial pressure. However, only a small number actually have an increase in arterial pressure to a level>140/90, and, in about half of these, the hypertension will remit within 6 months of stopping the drug.

Why some women taking oral contraceptives develop hypertension and others do not is unclear but may be related to (1) increased vascular sensitivity to angiotensin II, (2) the presence of mild renal disease, (3) familial factors (over one-half have a positive family history for hypertension), (4) age (hypertension is significantly more prevalent in women over age 35), (5) the estrogen content of the contraceptive, and/or (6) obesity. Indeed some investigators have suggested that the oral contraceptives are simply unmasking women with essential hypertension.

COARCTATION OF THE AORTA (See also Chap. 234)

The hypertension associated with coarctation may be caused by the constriction itself or perhaps by the changes in the renal circulation, which result in an unusual form of renal arterial hypertension. The diagnosis of coarctation is usually evident from physical examination and routine x-ray findings.

EFFECTS OF HYPERTENSION

Patients with hypertension die prematurely; the most common cause of death is heart

disease, with stroke and renal failure also frequent, particularly in patients with significant retinopathy.

EFFECTS ON THE HEART

Cardiac compensation for the excessive workload imposed by increased systemic pressure is at first sustained by concentric left ventricular hypertrophy, characterized by an increase in wall thickness. Ultimately, the function of this chamber deteriorates, the cavity dilates, and the symptoms and signs of heart failure appear (Chap. 231). Angina pectoris may also occur because of the combination of accelerated coronary arterial disease and increased myocardial oxygen requirements as a consequence of the increased myocardial mass (Chap. 244). On physical examination, the heart is enlarged and has a prominent left ventricular impulse. The sound of aortic closure is accentuated, and there may be a faint murmur of aortic regurgitation. Presystolic (atrial, fourth) heart sounds appear frequently in hypertensive heart disease, and a protodiastolic (ventricular, third) heart sound or summation gallop rhythm may be present. Electrocardiographic changes of left ventricular hypertrophy (Chap. 226) may occur, but the electrocardiogram substantially underestimates the frequency of cardiac hypertrophy compared with that observed with the echocardiogram. Evidence of ischemia or infarction may be observed late in the disease. Most deaths due to hypertension result from myocardial infarction or congestive heart failure. Recent data suggest that some of the myocardial damage may be mediated by aldosterone in the presence of a normal/high salt intake rather than just the increased blood pressure or an increase in angiotensin II levels per se.

NEUROLOGIC EFFECTS

The neurologic effects of long-standing hypertension may be divided into retinal and central nervous system changes. Because the retina is the only tissue in which the arteries and arterioles can be examined directly, repeated ophthalmoscopic examination provides the opportunity to observe the progress of the vascular effects of hypertension (Table 35-2). The Keith-Wagener-Barker classification of the *retinal changes* in hypertension has provided a simple and excellent means for serial evaluation of hypertensive patients. Increasing severity of hypertension is associated with focal spasm and progressive general narrowing of the arterioles, as well as the appearance of hemorrhages, exudates, and papilledema. These retinal lesions often produce scotomata, blurred vision, and even blindness, especially when there is papilledema or hemorrhages of the macular area. Hypertensive lesions may develop acutely and, if therapy results in significant reduction of blood pressure, may show rapid resolution. Rarely, these lesions resolve without therapy. In contrast, retinal arteriolosclerosis results from endothelial and muscular proliferation, and it accurately reflects similar changes in other organs. Sclerotic changes do not develop as rapidly as hypertensive lesions, nor do they regress appreciably with therapy. As a consequence of increased wall thickness and rigidity, sclerotic arterioles distort and compress the veins where the two vessel types cross in their common fibrous sheath, and the reflected light streak from the arterioles is changed by the increased opacity of the vessel wall.

Central nervous system dysfunction also occurs frequently in patients with hypertension. Occipital headaches, most often occurring in the morning, are among the most

prominent early symptoms of hypertension. Dizziness, light-headedness, vertigo, tinnitus, and dimmed vision or syncope may also be observed, but the more serious manifestations are due to vascular occlusion, hemorrhage, or encephalopathy (Chap. 361). The pathogeneses of the former two disorders are quite different. *Cerebral infarction* is secondary to the increased atherosclerosis observed in hypertensive patients, whereas *cerebral hemorrhage* is the result of both the elevated arterial pressure and the development of cerebral vascular microaneurysms (Charcot-Bouchard aneurysms). Only age and arterial pressure are known to influence the development of the microaneurysms. Thus, it is not surprising that arterial pressure shows a better association with cerebral hemorrhage than with either cerebral or myocardial infarction.

Hypertensive encephalopathy consists of the following symptom complex: severe hypertension, disordered consciousness, increased intracranial pressure, retinopathy with papilledema, and seizures. The pathogenesis is uncertain but is probably not related to arteriolar spasm or cerebral edema. Focal neurologic signs are infrequent and, if present, suggest that infarction, hemorrhage, or transient ischemic attacks are more likely diagnoses. Although some investigators have suggested that prompt lowering of arterial pressure in these patients may adversely affect cerebral blood flow, most studies indicate that this is not the case.

EFFECTS ON THE KIDNEY (See alsoChap. 278)

Arteriosclerotic lesions of the afferent and efferent arterioles and the glomerular capillary tufts are the most common renal vascular lesions in hypertension and result in a decreased glomerular filtration rate and tubular dysfunction. Proteinuria and microscopic hematuria occur because of glomerular lesions, and approximately 10% of the deaths caused by hypertension result from renal failure. Blood loss in hypertension occurs not only from renal lesions; epistaxis, hemoptysis, and metrorrhagia also occur frequently in these patients.

Approach to the Patient

The detailed initial evaluation of the hypertensive patient is outlined inChap. 35. It includes the critical elements of the history, physical examination, and basic laboratory investigation that aid in arriving at appropriate diagnostic and therapeutic decisions (Table 35-2).

DIAGNOSIS OF SECONDARY HYPERTENSION

Certain clues from the history, physical examination, and basic laboratory studies may suggest an unusual cause for the hypertension and dictate the need for special studies. For example, the abrupt onset of severe hypertension and/or the onset of hypertension of any severity in a patient under the age of 25 or over the age of 50 should lead to laboratory tests to exclude renovascular hypertension and pheochromocytoma. A history of headaches, palpitations, anxiety attacks, unusual sweating, hyperglycemia, and weight loss should also lead to tests to exclude pheochromocytoma. The presence of an abdominal bruit should lead to a workup for renovascular hypertension, and the finding on physical examination of bilateral upper abdominal masses consistent with polycystic renal disease should lead to the performance of an abdominal ultrasound

examination or intravenous pyelogram (IVP). An elevated creatinine or blood urea nitrogen level, associated with proteinuria and hematuria, should prompt a detailed workup for renal insufficiency (Chap. 268). Special studies for secondary hypertension are also indicated if there is therapeutic failure with the initial drug program. The specific diagnostic measures depend on the most likely causes of secondary hypertension.

Pheochromocytoma (See also Chap. 332) The easiest and best screening procedure for pheochromocytoma is the measurement of catecholamines and their metabolites in a 24-h urine sample collected while the patient is hypertensive. Measurement of plasma catecholamine levels may also be useful. These tests may be indicated even in patients who do not have episodic hypertension, since over half the patients with pheochromocytoma have fixed hypertension. Provocative tests are seldom, if ever, indicated, although occasionally a suppressive test may be useful.

Cushing's Syndrome (See alsoChap. 331) A 24-h urine test for cortisol and creatinine or the administration of 1 mg of dexamethasone at bedtime, followed by the measurement of plasma cortisol at 7 to 10 A.M., is the best test to screen for the presence of Cushing's syndrome. A urine cortisol level of<2750 nmol (100 ug) or suppression of the plasma cortisol level to <140 nmol/L (5 ug/dL) effectively rules out Cushing's syndrome.

Renovascular Hypertension (See also Chap. 278) Over the past decades the standard approach to screen for renovascular hypertension has progressed from the rapid-sequence IVP to one of three noninvasive techniques: the captopril-enhanced radionuclide renal scan (the preferred choice), a duplex Doppler flow study, or magnetic resonance (MRI) angiography. However, perhaps the most sensitive and specific screening test, the spiral computed tomography (CT) scan, which gives a three-dimensional view, unfortunately also requires giving an intravenous contrast agent.

The definitive test for surgically correctable renal disease is the combination of a renal angiogram and renal vein renin determinations. The renal arteriogram both establishes the presence of a renal arterial lesion and aids in the determination of whether the lesion is due to atherosclerosis or to one of the fibrous or fibromuscular dysplasias. It does not, however, prove that the lesion is responsible for the hypertension, nor does it permit prediction of the chances of surgical cure. It must be noted that (1) renal artery stenosis is a frequent finding by angiography and at postmortem in normotensive individuals, and (2) essential hypertension is a common condition and may occur in combination with renal arterial stenosis that is not responsible for the hypertension. Bilateral renal vein catheterization for measurement of plasma renin activity is therefore used to assess the functional significance of any lesion noted on arteriography. When one kidney is ischemic and the other is normal, all the renin released comes from the involved kidney. In the most straightforward situation, the ischemic kidney has a significantly higher venous plasma renin activity than the normal kidney, by a factor of 1.5 or more. Moreover, the renal venous blood draining the uninvolved kidney exhibits levels similar to those in the inferior vena cava below the entrance of the renal veins.

Significant benefit from operative correction may be anticipated in at least 80% of patients with the findings described above if care is taken to prepare the patient properly

before renal vein blood sampling, i.e., by discontinuing renin-suppressing drugs, such as beta blockers, for at least 10 days; restricting the patient to a low-sodium intake for 4 days; and/or giving a converting-enzyme inhibitor for 24 h. When obstructing lesions in the *branches* of the renal arteries are demonstrated by arteriography, an attempt to obtain blood samples from the main *branches* of the renal vein should be made in an effort to identify a localized intrarenal arterial lesion responsible for the hypertension.

Primary Aldosteronism (See also Chap. 331) These patients usually exhibit hypokalemia. Diuretic therapy often complicates the picture when the hypokalemia is first observed and needs to be assessed. Given the presence of hypokalemia, the relation between plasma renin activity and the aldosterone level becomes the key to the diagnosis of primary aldosteronism. The aldosterone concentration or excretion rate is high and plasma renin activity is low in primary aldosteronism, and these levels are relatively unaffected by changes in sodium balance. Thus, the aldosterone:renin ratio is high. A critical part of the evaluation after primary aldosteronism has been established is to determine whether disease is unilateral or bilateral, because surgical removal of the lesion usually reduces arterial pressure only in patients with unilateral disease.

Plasma Renin Activity Measurements Some studies have suggested that the plasma renin level should be measured in most hypertensive patients and related to a 24-h urine sodium excretion rate to assess whether high, low, or normal renin levels are present. It has been proposed that this information may be important for both therapeutic and prognostic reasons. However, as noted earlier, it is unclear, on the basis of the available data and treatment programs, that these random measurements are really useful except in patients with findings suggestive of renal vascular disease or mineralocorticoid excess in whom lateralizing renal vein renin levels or suppressed peripheral renin levels may be of diagnostic and/or therapeutic significance.

TREATMENT

Indications for Therapy Virtually every patient with a diastolic arterial pressure that persistently exceeds 90 mmHg, or any patient over 65 years of age with a systolic arterial pressure >160 mmHg, is a candidate for diagnostic studies and for subsequent treatment. Furthermore, at any given level of blood pressure elevation, the ultimate risk of developing hypertensive vascular complications is greater in men than in women, in younger than in older persons, and in diabetic than nondiabetic patients. It may be argued, then, that it is hard to justify producing the uncomfortable side effects of therapy in, for example, an asymptomatic woman over 70 years of age with a diastolic pressure of 90 mmHg. On the other hand, it is easy to justify side effects in a man of 30 with a diastolic pressure exceeding 110 mmHg because such a person may be expected to receive the greatest benefit from therapy. Fortunately, the choice of treatment is such that a satisfactory program to control arterial pressure with minimal side effects can be developed for most patients, particularly as more studies assessing the impact of specific therapeutic agents on the patient's quality of life are reported.

A reasonable guideline would be that all patients with a diastolic pressure repeatedly>90 mmHg or systolic pressure>140 mmHg should be treated unless specific contraindications exist. Patients with isolated *systolic* hypertension (levels >160 mmHg with diastolic pressure<89 mmHg) should also be treated if they are over age 65. It is

uncertain that individuals under age 65 who have isolated systolic hypertension will benefit from therapy until the results of a well-controlled, prospective study are completed. Patients with labile hypertension or isolated systolic hypertension who are not treated should have regular follow-up examinations at 6-month intervals because of the frequent development of progressive and/or sustained hypertension. Finally, if coronary artery disease or associated cardiovascular risks are present, then treatment of a patient with a lower blood pressure may be warranted. For example, patients with angina pectoris or diabetes mellitus with diastolic blood pressures between 85 and 90 mmHg may be candidates for antihypertensive therapy.

What should the blood pressure goal be? Previously it was assumed that 140/90 mmHg was the desired level. This still seems reasonable for nondiabetic patients since the Hypertension Optimal Treatment (HOT) study did not detect a significant difference in cardiovascular risk between patients with treatment goal diastolic blood pressures of 90 and 80 mmHg. However, in patients with diabetes this is not the case. In the UK Prospective Diabetes Study (UKPDS), individuals with a blood pressure of 144/82 mmHg had a substantially lower risk compared to those with a blood pressure of 154/87 mmHg. The HOT study investigators documented a similar finding in their diabetic subset. Thus, it seems reasonable to target a blood pressure in the normal range for diabetic patients, i.e., 130/85 mmHg. While not definitively proven, it seems prudent to use the same goal in all young and middle-aged patients depending on what other cardiovascular risk factors are present. For elderly individuals, a goal of 140/90 mmHg is appropriate, although definitive data for lowering systolic blood pressure below 160 mmHg is still lacking. Importantly, how aggressive one should be in achieving these blood pressure goals depends on the number and severity of other risk factors present.

The identification of an operable form of secondary hypertension does not automatically mean that surgical treatment is indicated. The decision depends on the age and general health of the patient, the natural history of the lesion, and the response of the arterial pressure to drug therapy. In patients with renovascular hypertension, the feasibility of renal angioplasty, the advantages of surgical repair versus nephrectomy, and the degree of overall renal functional impairment must be considered. Age and general health are important in patients with renovascular hypertension due to arteriosclerosis, because there is no evidence that repair of the stenosis increases life expectancy in the elderly patient with other evidence of vascular disease. Knowledge of the natural history of the disease is especially important when making a decision in the case of a young patient with renal artery stenosis due to fibrous dysplasia. If the arteriographic appearance suggests that the stenosis is due to intimal or subadventitial fibroplasia, the lesion may be expected to progress, and operation or angioplasty is required. Medial fibroplasia, on the other hand, often remains stable, and operation or angioplasty may not be necessary if pressure can be controlled by drug therapy.

The decision regarding operation should also be considered carefully in patients with primary aldosteronism when neither abdominalCT nor bilateral adrenal venous sampling for aldosterone demonstrates a tumor, because such patients may prove to have multinodular hyperplasia. In that case, bilateral adrenalectomy would be required to eliminate the aldosterone excess, and, even then, hypertension would usually persist. If hypokalemia can be controlled by an aldosterone receptor antagonist, e.g., spironolactone, or other drug therapy and arterial pressure lowered with

antihypertensive agents, then it is reasonable to withhold operative treatment.

GENERAL MEASURES

Nondrug therapeutic intervention is probably indicated in all patients with sustained hypertension and probably in most with labile hypertension. The general measures employed include (1) relief of stress, (2) dietary management, (3) regular aerobic exercise, (4) weight reduction (if needed), and (5) control of other risk factors contributing to the development of arteriosclerosis. Relief of emotional and environmental stress is one of the reasons for the improvement in hypertension that occurs when a patient is hospitalized. Though it is usually impossible to extricate the hypertensive patient from all internal and external stresses, he or she should be advised to avoid unnecessary tensions. In rare instances, it may be appropriate to recommend a change of job or of life-style. It has been suggested that relaxation techniques may also lower arterial pressure. However, it is uncertain that these techniques alone have much long-term effect.

Dietary management has three aspects:

1. Because of the documented efficacy of sodium restriction and volume contraction in lowering blood pressure, patients previously were instructed to curtail sodium intake drastically. Some investigators have suggested that this is not necessary. They base their conclusion on two observations: (1) In many patients the blood pressure is not sensitive to the level of sodium intake, and (2) diuretics provide another method of decreasing body sodium stores in individuals whose blood pressure is sodium-sensitive. However, meta-analyses of previous diet studies have documented a 5-mmHg reduction in systolic pressure and a 2.6-mmHg reduction in diastolic pressure when sodium intake is reduced by approximately 75 meq/d. In addition, several reports have documented that, while mild sodium restriction has little if any direct action on blood pressure, it significantly potentiates the efficacy of nearly all antihypertensive agents. Thus, by making it possible to control blood pressure with lower doses of drugs, sodium restriction leads to a reduction in side effects. In addition, it is quite clear that in some hypertensive patients, as noted above, the level of sodium intake does influence the blood pressure. Thus, since there is no apparent risk to mild sodium restriction, the most practical approach now is to advise mild dietary sodium restriction (up to 5 g NaCl per day), which can be achieved by eliminating all additions of salt to food that is prepared normally. Some studies have also reported a lowering of arterial pressure related to an *increase* in potassium and/or calcium intake. For example, in one meta-analysis, dietary potassium supplements of 50 to 120 meq/d reduced blood pressure by about the same amount as salt restriction (by 6 mmHg systolic and 3.4 mmHg diastolic). While the advisability of these forms of dietary alteration is still controversial, the fact that a moderately high calcium intake (1.5 g elemental calcium daily) probably also reduces the extent of age-related osteoporosis, combined with the results of the potassium supplementation studies, indicate that they are probably useful adjuncts. A particularly useful approach is the DASH (Dietary Approaches to Stop Hypertension) diet, which uses natural foods that are high in potassium and low in saturated and total fat. This diet significantly lowered blood pressure in borderline and stage 1 hypertensive subjects (see Table 35-1 for definitions).

2. Caloric restriction should be urged for patients who are overweight. Some obese patients will show a significant reduction in blood pressure simply as a consequence of weight loss. In the Trial of Antihypertensive Interventions and Management (TAIM) study, weight reduction (average 4.4 kg over 6 months) lowered blood pressure by 2.5 mmHg.

3. A restriction in the intake of cholesterol and saturated fats is recommended, as this diet modification may diminish the incidence of arteriosclerotic complications. Reducing alcohol intake to <15 mL daily is also beneficial. Regular exercise is indicated within the limits of the patient's cardiovascular status. Not only is exercise helpful in controlling weight, but there is also evidence that physical conditioning itself may lower arterial pressure. Isotonic exercises (jogging, swimming) are better than isometric exercises (weight lifting) since the latter, if anything, raises arterial pressure. The dietary management outlined above is aimed at the control of other risk factors. Probably the most significant additional step that could be taken in this area would be to convince the smoker to give up cigarettes.

DRUG THERAPY FOR HYPERTENSION (Table 246-6)

To make rational use of antihypertensive drugs, the sites and mechanisms of their action must be understood. In general, there are six classes of drugs: diuretics, antiadrenergic agents, vasodilators, calcium entry blockers, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor antagonists.

DIURETICS (See alsoChap. 232)

The thiazides are the most frequently used and most extensively investigated members of this group, and their early effect is certainly related to sodium diuresis and volume depletion. A reduction in peripheral vascular resistance has also been reported by some workers to be important in the long term. Traditionally, thiazide diuretics have formed the cornerstone of most therapeutic programs designed to lower arterial pressure, and they are usually effective within 3 to 4 days. Furthermore, they have been shown to reduce mortality and morbidity in long-term trials. However, in recent years there has been increasing resistance to their routine use, primarily because of their adverse metabolic effects, which include hypokalemia due to renal potassium loss, hyperuricemia due to uric acid retention, carbohydrate intolerance, and hyperlipidemia. These effects are minimized if the dose is kept below the equivalent of 25 mg/d of hydrochlorothiazide. The more potent loop-acting diuretics furosemide and bumetanide have also been shown to be antihypertensive but have been used less extensively for this indication, primarily because of their shorter duration of action. Spironolactone causes renal sodium loss by blocking the effect of mineralocorticoids, and, therefore, it may be more effective in patients whose mineralocorticoid levels are excessive, such as patients with primary or secondary aldosteronism. However, a clinical trial in heart failure using low doses of spironolactone achieved a 30% reduction in mortality, suggesting that an aldosterone receptor antagonist may be beneficial even when aldosterone levels are relatively normal. Although they do not compete directly with aldosterone, triamterene and amiloride act at the same site as spironolactone to impede sodium reabsorption. They are effective in the same situations as an aldosterone receptor antagonist, except that triamterene has little intrinsic antihypertensive effect.

Their major disadvantage is that they can produce hyperkalemia, particularly in patients with impaired renal function. Any of these three potassium-sparing diuretics can also be given along with thiazide diuretics to minimize renal potassium loss.

ANTIADRENERGIC AGENTS (See alsoChap. 72)

These drugs act at one or more sites -- centrally on the vasomotor center, in peripheral neurons, where they modify catecholamine release, or in target tissues, where they block adrenergic receptor sites. Drugs that appear to have predominant *central actions* are *clonidine*, *methyldopa*, *guanabenz*, and *guanfacine*. These drugs and their metabolites are predominantly a-receptor agonists. Stimulation of a_2 receptors in the vasomotor centers of the brain *reduces* sympathetic outflow, thereby reducing arterial pressure. Usually a fall in cardiac output and heart rate also occurs, more commonly with clonidine and guanabenz, but the baroreceptor reflex is intact. Thus, postural symptoms are absent. However, rebound hypertension may occur rarely when these drugs, particularly clonidine and guanabenz, are stopped. This effect is probably secondary to an increase in norepinephrine release, which is inhibited by these agents owing their agonist effect on presynaptic a receptors. They are usually not used as first-line therapy.

Another class of antiadrenergic agents consists of the *ganglionic blocking drugs*, which are used infrequently now. Because of their side effects, ganglionic blocking agents are now usually reserved for the rapid lowering of arterial pressure by parenteral administration of the short-acting agent *trimethaphan* in patients with severe hypertension.

Various drugs act at *postganglionic adrenergic nerve endings*, but they are rarely used now because of their side effects. *Guanethidine* and its shorter-acting analogue guanadrel block the release of norepinephrine from adrenergic nerve endings. They usually reduce cardiac output and lower systolic more than diastolic blood pressure. They also produce a greater postural effect than the other drugs that act at the nerve endings, and orthostatic hypotension is a frequent side effect.

The last group of drugs affecting the adrenergic system are those that block the *peripheral adrenergic receptors*, a,b, or both (Chap. 72).

a-Adrenergic Receptor Blockers These agents also usually are not used as first-line therapy. *Phentolamine* and *phenoxybenzamine* block the action of norepinephrine at a-adrenergic receptor sites. These two compounds block both presynaptic (a_2) and postsynaptic (a_1) a receptors, and the former action accounts for the tolerance that develops. *Prazosin* is more effective because it selectively blocks only *postsynaptic* a receptors, i.e., a_1 receptors. Thus, presynaptic a activity remains, suppressing norepinephrine release, and tolerance occurs only infrequently. Accordingly, prazosin produces less tachycardia but more postural hypotension than direct-acting vasodilators, such as hydralazine, and rarely can produce substantial hypotension following the first dose. Its use has decreased with a report of its association with an increase in cardiovascular events.

b-Adrenergic Receptor Blockers (See also Chap. 244) A number of

effective*b-adrenergic receptor blocking agents* are available that block sympathetic effects on the heart and should be most effective in reducing cardiac output and in lowering arterial pressure when there is increased cardiac sympathetic nerve activity. These agents *are* often used as first-line therapy. In addition, they block the adrenergic nerve-mediated release of renin from the renal juxtaglomerular cells. This action may be an important component of their blood pressure-lowering action.b-Adrenergic blockers are particularly useful when employed in conjunction with vascular smooth-muscle relaxants, which tend to evoke a reflex increase in heart rate, and with diuretics, the administration of which often results in an elevation of circulating renin activity. In practice, beta blockers appear to be effective even when there is no evidence of increased sympathetic tone, with about one-half or more of all hypertensive patients showing a fall in pressure. Furthermore, like diuretics, they have been shown to reduce morbidity and mortality in long-term clinical trials. However, these agents can precipitate congestive heart failure and asthma in susceptible individuals, and they must be used with caution in diabetic patients receiving hypoglycemic therapy because they inhibit the usual sympathetic responses to hypoglycemia. Cardioselective beta-blocking agents (so-called $beta_1$blockers, e.g., metoprolol, atenolol) have been developed and may be superior to nonselective beta blockers such as propranolol and timolol in patients with bronchospasm. Nadolol, a nonselective beta blocker, unlike other drugs of this class, is excreted unchanged in the urine and has a half-life of 14 to 20 h; only one dose a day is required. Atenolol also usually needs to be given only once a day. Pindolol and acebutolol are nonselective beta blockers that have partial agonist activity and, therefore, produce less bradycardia. Labetalol exerts both a- andb-adrenergic blocking actions. It is usually not used as first-line therapy as there is no mortality study in which it has been tested. Thus, it lowers arterial pressure not only by the same complex actions as do beta blockers but also directly by reducing systemic vascular resistance. Usually it has a more rapid onset of action but produces more postural symptoms and chronic sexual dysfunction than the other beta blockers.

VASODILATORS

These agents are usually not used for initial therapy. *Hydralazine* is the most versatile of the drugs that cause direct relaxation of vascular smooth muscle; it is effective both orally and parenterally and acts mainly on arterial resistance rather than on venous capacitance vessels, as evidenced by lack of postural effects. Unfortunately, the effect of hydralazine on peripheral resistance is partly negated by a reflex increase in sympathetic discharge that raises heart rate and cardiac output. This response limits the usefulness of hydralazine, especially in patients with severe coronary artery disease. However, the efficacy of hydralazine can be increased if it is given in conjunction with a beta blocker or a drug such as methyldopa or clonidine, all of which block reflex sympathetic stimulation of the heart. A serious side effect of doses of hydralazine exceeding 300 mg/d has been the production of a lupus erythematosus-like syndrome.

Minoxidil is even more potent than hydralazine but unfortunately produces significant hypertrichosis and fluid retention and, therefore, is mainly limited to patients with severe hypertension and renal insufficiency.

Diazoxide, a thiazide derivative, is restricted in its application to acute situations. It is not a diuretic; in fact, it causes sodium retention. However, like other thiazides, it reduces

carbohydrate tolerance. It must be given rapidly intravenously to guarantee an effect. It begins to act immediately to lower blood pressure, and its effects may last for several hours. *Nitroprusside* given intravenously also acts as a direct vasodilator, with onset and offset of actions that are almost immediate. *Nitroglycerin* is a third direct-acting vasodilator useful as an intravenous agent. These latter three drugs are useful only for the treatment of hypertensive emergencies (Table 246-7).

ACE INHIBITORS

Drugs from several of the categories discussed above have been shown to possess an additional action resulting in inhibition of renin secretion. These include clonidine, reserpine, methyldopa, and beta blockers. A second group of drugs inhibit the enzyme converting angiotensin I into angiotensin II -- ACE. These agents are an increasingly popular choice for initial therapy. They are useful because they not only inhibit the generation of a potent vasoconstrictor (angiotensin II) but also may retard the degradation of a potent vasodilator (bradykinin), alter prostaglandin production (an effect most notable with captopril), and can modify the activity of the adrenergic nervous system. They are especially useful in renal or renovascular hypertension and in diabetic patients, as well as in accelerated and malignant hypertension. However, in patients with bilateral renal artery stenosis, rapid deterioration of renal function may occur. They are also as effective in mild, uncomplicated hypertension as beta blockers and thiazides -- and have fewer side effects, particularly ones that adversely affect the patient's quality of life.

These drugs should be used with caution when the renin system is activated (e.g., by severe heart failure, prior diuretic therapy, or substantial salt restriction) to avoid profound hypotension. Usually, diuretics are stopped 2 to 3 days before administration of an ACEinhibitor is begun and are added back later if needed.

ANGIOTENSIN RECEPTOR ANTAGONISTS

These drugs have effects similar to those ofACEinhibitors. However, instead of blocking the production of angiotensin II, they competitively inhibit its binding to the angiotensin II AT_1receptor subtype. Their utility and tolerability are similar to those of the ACE inhibitors, but they do not cause cough or angioedema.

CALCIUM CHANNEL ANTAGONISTS

There are three subclasses of calcium channel antagonists: the phenylalkylamine derivatives (e.g., verapamil), the benzothiazepines (e.g., diltiazem), and the dihydropyridines (e.g., amlodipine). To date, there is only one therapeutic agent in each of the first two classes but a number of agents in the third class. All three subclasses modify calcium entry into cells by interacting with specific binding sites on the$\alpha_1$subunit of the L-type voltage-dependent calcium channel. Thus, since there are other calcium channels (e.g., the T and N types), the actions of these drugs only partially modify total calcium transport into cells. The relative specificity of each agent stems from the fact that each class has a unique binding site on the$\alpha_1$subunit, and these sites are variably expressed in different tissues. Thus, while agents from all three subclasses cause vasodilation, usually only dihydropyridines produce reflex tachycardia. Diltiazem and

verapamil can both slow atrioventricular conduction -- a feature not observed with the dihydropyridines. While calcium channel antagonists are also useful in angina pectoris (Chap. 244), because of their negative inotropic actions, they should be used with caution in hypertensive patients with heart failure. Considerable controversy has surrounded the use of calcium channel antagonists in the treatment of hypertension. In part the controversy was secondary to the inadequacy of the data and the confusion between the use of short-acting agents (e.g., nifedipine) and long-acting agents. Several facts have helped partially to resolve this controversy. First has been the general recognition that despite its previously frequent use as an antihypertensive agent, short-acting nifedipine rarely, if ever, should be used to treat hypertension, since it has been reported to increase the incidence of acute coronary events. Second, the results of the SYST-EUR (Systolic Hypertension in Europe) trial documented that a long-acting calcium channel antagonist reduced mortality to an extent equivalent to that previously reported for diuretics and beta blockers. Thus, long-acting calcium channel antagonists are often used as first-line therapy.

APPROACH TO DRUG THERAPY

The aim of drug therapy is to use the agents just described, alone or in combination, to return arterial pressure to normal levels with minimal side effects. Ideally, one would choose a therapeutic program that specifically corrects the underlying defect resulting in the elevated blood pressure -- for example, treatment with spironolactone for patients with primary aldosteronism. As our knowledge of the mechanisms underlying the hypertension in individual patients increases, more specific drug programs will become available. Such programs presumably will result in normalization of blood pressure with fewer side effects. In the absence of this information, an empirical approach is used, which takes into consideration efficacy, safety, impact on the quality of life, compliance, ease of administration, and cost. When used in combination, drugs are chosen for their different sites of action. However, except for those patients with severe hypertension (average diastolic blood pressure>130 mmHg), in whom intensive therapy with several agents simultaneously is usually required, most patients are treated *initially* with a single agent.

Since many effective antihypertensive agents are available, a number of useful therapeutic regimens have been developed. There are two major authoritative groups who have treatment guidelines when the patient's condition does not require a specific approach: World Health Organization-International Society of Hypertension (WHO-ISH) (Figs. 246-1 and246-2) and the Sixth U.S. Joint National Committee (JNC) on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VI). In the absence of specific therapy, their approaches are similar in most respects, relying heavily on the results of randomized clinical trials (Table 246-8), except for which drugs should be used to initiate therapy. JNC VI recommends starting with diuretics and/or beta blockers because they are the ones where mortality trials have demonstrated a positive effect of treatment. The WHO-ISH guidelines recommends initiating therapy with any of six classes of agents (Table 246-9). The different recommendations, in part, reflect the fact that the WHO-ISH committee reviewed more recent data from mortality clinical trials that, in one case, documented a reduction in morbidity and mortality with a long-acting calcium channel antagonist versus placebo that was similar to previous reports for diuretics and beta blockers and, in another case, reported thatACEinhibitors

were as effective as beta blockers and diuretics in reducing mortality (Table 246-8).

There are several critical caveats common to both approaches:

1. Start with a low dose of an agent and, if blood pressure is not controlled, increase only moderately.

2. Start with an agent that may also treat and/or not harm a coexisting condition.

3. Add a second agent from a different, complementary class if blood pressure is not controlled with a moderate dose of the first agent.

4. Start with an agent that the patient is likely to tolerate best; long-term compliance is related to tolerability and efficacy of the first agent used.

5. Use a diuretic when two agents are used, in nearly all cases.

6. Use thiazide diuretics only at low doses, i.e., £25 mg/d of hydrochlorothiazide or its equivalent, unless some pressing reason exists.

7. Use low-dose combination therapy when appropriate as initial therapy:

a. A diuretic with a beta blocker, ACE inhibitor, or angiotensin II antagonist;

b. A calcium channel blocker with an ACE inhibitor or a beta blocker

8. One or two agents will control blood pressure in 90% of hypertensive patients; to achieve a diastolic blood pressure of <90 mmHg in the HOTstudy, two agents were required in 70% of cases.

If therapy with two drugs does not achieve blood pressure control, the primary agent should be increased to full dose, e.g., 100 mg of captopril or atenolol, 20 mg of enalapril, or 360 mg of diltiazem. If the blood pressure is still not controlled, then a detailed search for a secondary cause of hypertension, as outlined above, is indicated. If none is found, then a dietary assessment will often reveal a high sodium intake. With reduction in salt intake to 5 g/d or less, blood pressure is often controlled. If the blood pressure is still not controlled, then the primary agent should be switched, maintaining the thiazide. Caution should be used if anACEinhibitor was not the original agent, as administration of such an agent to a patient who is already taking a diuretic occasionally may lead to profound hypotension. If none of these changes produces better control of arterial pressure, then the combination of a calcium channel antagonist and an ACE inhibitor, or triple therapy, usually with a diuretic, ACE inhibitor, and hydralazine, may be effective.

If the blood pressure is controlled, then a stepwise reduction in the dose and/or withdrawal of some of the agents should be carried out to determine the minimal therapeutic program that will maintain the blood pressure at 140/90 mmHg or less. Whether triple or quadruple drug therapy is warranted to lower blood pressure further is uncertain.

Fewer than 5% of patients will still be hypertensive at this point. For these, one first should consider the reasons for therapeutic failure, as shown inTable 246-10. If none can be identified, then one of the other agents, such as one of the vasodilators listed inTable 246-6 (e.g., hydralazine) or an antiadrenergic agent (e.g., prazosin or clonidine), should be added. If blood pressure is controlled, previous drugs are withdrawn sequentially to determine the minimal therapeutic program that will maintain a normal blood pressure.

While the recommendations outlined above are satisfactory for a large majority of patients, it is important to use a flexible approach, because individual patients may respond differently to individual drugs and drug combinations. For those patients requiring multiple drugs, once the appropriate combination has been found, the use of a single formulation with the appropriate combination of drugs may simplify the regimen and thereby increase compliance. Every effort should be made to reduce the number of times each day the patients must interrupt their schedules for the medication. Pharmacologic treatment of essential hypertension is usually lifelong, and since most patients are asymptomatic, compliance with a complex regimen may be a serious problem, particularly if the therapeutic regimen has a negative impact on the quality of the patient's life. Finally, it is uncertain what level of arterial pressure should be accepted as representing adequate control. It is clear that reducing diastolic blood pressure to<90 mmHg is appropriate and reduces morbidity and/or mortality.

Five groups of patients with hypertension require special consideration because of associated conditions. These groups are considered in the following sections.

RENAL DISEASE

Reduction of arterial pressure in hypertensive patients with impaired renal function is often accompanied initially by an increase in serum creatinine. This change does not represent further structural renal damage and should not deter the physician from continuing the therapy, since achievement of blood pressure control may eventually reduce the value toward normal. However, if serum creatinine increases in a patient treated with a converting-enzyme inhibitor, care needs to be exercised, because these patients may have bilateral renal artery disease. Their renal function will continue to deteriorate as long as the converting-enzyme inhibitor is given. Thus, converting-enzyme inhibitors should be used cautiously in patients with impaired renal function, and renal function should be assessed frequently (every 4 to 5 days) for the first 3 weeks. While converting-enzyme inhibitors are contraindicated in patients with bilateral renal artery stenosis, these are the drugs of choice in patients with unilateral renal artery stenosis and a normally functioning contralateral kidney and probably also in patients with chronic renal failure with or without diabetes mellitus.

CORONARY ARTERY DISEASE

In these patients, who also may be taking cardiac glycosides, thiazides should be used judiciously, and a reduction in serum potassium levels should be watched for and, if found, should be corrected rapidly. Beta blockers should be withdrawn carefully, if at all, in these patients. Finally, calcium channel antagonists and converting-enzyme inhibitors

may be useful in these patients because they minimize a number of potential adverse reactions that accompany the use of other therapeutic agents, particularly nonspecific vasodilators.

DIABETES MELLITUS

The diabetic patient with hypertension is particularly challenging to treat because many of the agents used to lower blood pressure can affect glucose metabolism adversely. Converting-enzyme inhibitors may be particularly useful in these individuals. They have no known adverse effects on glucose or lipid metabolism and minimize the development of diabetic nephropathy by reducing renal vascular resistance and renal perfusion pressure -- the primary factor underlying renal deterioration in these patients.

PREGNANCY

The patient who is pregnant and hypertensive or who develops hypertension during pregnancy (pregnancy-induced hypertension, preeclampsia, eclampsia) is particularly difficult to treat. Because it is uncertain whether autoregulation of uterine blood flow occurs, lowering blood pressure in the pregnant hypertensive patient may result in reduced placental and fetal perfusion. Thus, a conservative approach to lowering blood pressure is usually indicated. In the second and third trimesters, antihypertensive agents are often not indicated unless the diastolic pressure exceeds 95 mmHg. In general, severe salt restriction and/or diuretics are not used because of the associated increase in fetal wastage. Beta blockers need to be used cautiously for similar reasons. Methyldopa and hydralazine, and to a lesser extent calcium channel antagonists, are the antihypertensive agents used most often, because they have no known adverse effects on the fetus. Little is known about the safety of other antihypertensive agents in pregnancy, except that nitroprusside and converting-enzyme inhibitors may cause adverse effects on the fetus and are contraindicated.

ELDERLY PATIENTS

Hypertensive patients who are over age 65, and particularly those over age 75, offer substantial challenges to the physician. Several studies have reported that healthy elderly patients, whether male or female, who are treated with relatively modest doses of antihypertensive agents show a substantial reduction in strokes and stroke-related deaths. This is true whether the patient has systolic and diastolic hypertension or isolated systolic hypertension. What is not clear from these studies is how broadly the results can be extrapolated, since the studies were performed in healthy elderly patients, while many such patients have other diseases. Thus, in the elderly hypertensive patient, individualization of therapy still seems warranted.

Probably fewer than one-third of hypertensive patients in the United States are being treated effectively. Only a small number of these failures are related to drug unresponsiveness. Most are related to (1) failure to detect hypertension, (2) failure to institute effective treatment of an asymptomatic hypertensive patient, and (3) failure of the asymptomatic hypertensive patient to adhere to therapy. To help with the latter problem, patients must be educated to continue treatment once an effective regimen has been identified. Side effects and inconveniences of treatment must be minimized or

counteracted in order to obtain the patient's continued cooperation.

MALIGNANT HYPERTENSION

In addition to marked blood pressure elevation in association with papilledema and retinal hemorrhages and exudates, the full-blown picture of malignant hypertension may include manifestations of hypertensive encephalopathy, such as severe headache, vomiting, visual disturbances (including transient blindness), transient paralyses, convulsions, stupor, and coma. These manifestations have been attributed to spasm of cerebral vessels and to cerebral edema. In some patients who have died, multiple small thrombi have been found in the cerebral vessels. Cardiac decompensation and rapidly declining renal function are other critical features of malignant hypertension. Oliguria may, in fact, be the presenting feature. The vascular lesion characteristic of malignant hypertension is fibrinoid necrosis of the walls of small arteries and arterioles, and this development can be reversed by effective antihypertensive therapy.

The pathogenesis of malignant hypertension is unknown. However, at least two independent processes -- dilation of cerebral arteries and generalized arteriolar fibrinoid necrosis -- contribute to the associated signs and symptoms. The cerebral arteries dilate because the normal autoregulation of cerebral blood flow decompensates as a result of the markedly elevated arterial pressure. Cerebral blood flow therefore is excessive, producing the encephalopathy associated with malignant hypertension. Many patients also show evidence of a microangiopathic hemolytic anemia; this secondary phenomenon could contribute to the deterioration of renal function. Most patients also have elevated levels of peripheral plasma renin activity and increased aldosterone production, and these effects may be involved in causing vascular damage.

Perhaps fewer than 1% of hypertensive patients develop the malignant phase, which can occur in the course of both essential and secondary hypertension. Rarely, it is the first recognized manifestation of the blood pressure problem, and it is unusual for it to occur in patients under treatment. The average age at diagnosis is 40, and men are affected more often than women. Prior to the availability of effective therapy, the life expectancy after diagnosis of malignant hypertension was less than 2 years, with most deaths being due to renal failure, cerebral hemorrhage, or congestive heart failure. With the advent of effective antihypertensive therapy, at least half the patients survive for more than 5 years.

TREATMENT

Malignant hypertension is a medical emergency that requires immediate therapy. However, it needs to be distinguished from severe hypertension, since overly aggressive therapy in malignant hypertension could result in a potentially hazardous reduction in myocardial and cerebral perfusion. The initial aims of therapy should be (1) correction of medical complications and (2) reduction of diastolic pressure by one-third, but not to a level <95 mmHg. The drugs available for treatment of malignant hypertension can be divided into two groups on the basis of time of onset of action (Table 246-7). Those in the first group act within a few minutes but are not satisfactory for long-term management. If the patient is having convulsions, and if arterial pressure must be reduced rapidly, then one from the immediate-acting group should be used.

The first three agents in this group require continuous infusion and close monitoring. *Nitroprusside* is given by continuous intravenous infusion at a dose of 0.25 to 8.0 ug/kg per min. It is probably the agent of choice in this condition, since it dilates both arterioles and veins. It has the advantage over the ganglionic blockers of not being associated with the development of tachyphylaxis and can be used for days with few side effects. The dosage must be controlled with an infusion pump. *Nitroglycerin* affects veins more than arterioles and is given by continuous infusion at a rate of 5 to 100 ug/min. It is particularly useful in the treatment of hypertension following coronary bypass surgery, myocardial infarction, left ventricular failure, or unstable angina pectoris. *Diazoxide* is the easiest agent to administer, for no individual titration of dosage is required. However, it is probably less effective than the other agents. It primarily affects arteriolar and not venous tone. A dose of 50 to 100 mg is given rapidly intravenously, and the antihypertensive effect appears in 1 to 5 min. The same dose can be repeated in 5 to 10 min, if necessary, or when the pressure begins to rise, usually after several hours. The total dose should not exceed 600 mg/d. In an occasional patient, pressure may drop below normal levels after diazoxide administration. This drug should not be used in patients in whom aortic dissection or myocardial infarction is suspected. Because it can increase the force of myocardial contraction, often a beta blocker is given concomitantly. *Enalaprilat*, an intravenous form of theACEinhibitor *enalapril*, has also proven effective, particularly in individuals with left heart failure. Finally, intravenous *labetalol* may be particularly useful in patients with a myocardial infarct or angina because it prevents an increase in heart rate. However, it may be ineffective in patients previously treated with beta blockers and is contraindicated in patients with heart failure, asthma, bradycardia, or heart block. It may also serve as an alternative therapy in patients with eclampsia who are unresponsive to hydralazine.

Patients given any of these agents also should receive other medications effective for long-term control. Those in the second group in Table 246-7 require 30 min or more to produce their full effect, but they have the advantage of being satisfactory for subsequent oral administration and for long-term management of the patient's hypertension. If such a delay in the achievement of the full effect is acceptable, intravenous *hydralazine* is effective in many patients within 10 min; an effective protocol involves giving 10-mg doses intravenously every 10 to 15 min until the desired effect has been obtained or until a total of 50 mg has been administered. The total amount required for response may then be repeated intramuscularly or intravenously every 6 h. Hydralazine should be used with caution in patients with significant coronary artery disease and should be avoided in patients manifesting myocardial ischemia or aortic dissection. It is effective in preeclampsia. *Esmolol*, a beta blocker with an onset of action of 1 to 2 min, is particularly useful in aortic dissection and for perioperative hypertensive crisis. Its major disadvantage is that it can have a negative inotropic effect. Its use in individuals with congestive heart failure, obstructive lung disease, or asthma is problematic.

Furosemide is an important adjunct to the therapy just discussed. Given either orally or intravenously, it serves to maintain sodium diuresis in the face of a falling arterial pressure and thus will speed recovery from encephalopathy and congestive heart failure as well as maintain the sensitivity to the primary antihypertensive drug. Digitalis (Chap. 232) may also be indicated if there is evidence of cardiac decompensation.

In patients with malignant hypertension in whom the existence of pheochromocytoma is suspected, urine should be collected for measurement of the products of catecholamine metabolism, and drugs that might release additional catecholamines, such as methyldopa, reserpine, and guanethidine, must be avoided. The parenteral drug of choice in these patients is phentolamine, administered with care to avoid a precipitous reduction in arterial pressure.

There is hope even for patients who fail to respond sufficiently to any of the forms of therapy and who show progressive deterioration in renal function. In some, a period of peritoneal dialysis or hemodialysis to deplete extracellular fluid has resulted in better blood pressure control and eventual improvement in renal function. In other patients with refractory hypertension and renal failure who do not respond to volume depletion or hypotensive therapy, including minoxidil administration -- particularly those with marked elevation of plasma renin activity -- bilateral nephrectomy has resulted in amelioration of hypertension; subsequently, these patients have been maintained on chronic dialysis or have received renal homografts. However, bilateral nephrectomy should be avoided where possible because (1) the loss of renal erythropoietin will contribute to the associated anemia, (2) vitamin D metabolism may be adversely affected, and (3) all residual renal function will be lost.

(Bibliography omitted in Palm version)

Back to Table of Contents

247. DISEASES OF THE AORTA - *Victor J. Dzau, Mark A. Creager*

The aorta is the conduit through which the blood ejected from the left ventricle is delivered to the systemic arterial bed. In adults, its diameter is approximately 3 cm at the origin, 2.5 cm in the descending portion in the thorax, and 1.8 to 2 cm in the abdomen. The aortic wall consists of a thin intima composed of endothelium, subendothelial connective tissue, and an internal elastic lamina; a thick tunica media composed of smooth-muscle cells and extracellular matrix; and an adventitia composed primarily of connective tissue enclosing the vasa vasorum and nervi vascularis. In addition to its conduit function, the viscoelastic and compliant properties of the aorta also subserve a buffering function. The aorta is distended during systole to enable a portion of the stroke volume to be stored, and it recoils during diastole so that blood continues to flow to the periphery. Because of its continuous exposure to high pulsatile pressure and shear stress, the aorta is particularly prone to injury and disease resulting from mechanical trauma (Table 247-1). The aorta is also more prone to rupture than any other vessel, especially with the development of aneurysmal dilatation, since its wall tension, as governed by Laplace's law (i.e., proportional to the product of pressure and radius), would be increased.

AORTIC ANEURYSM

An *aneurysm* is defined as a pathologic dilatation of a segment of a blood vessel. A *true aneurysm* involves all three layers of the vessel wall and is distinguished from a *pseudoaneurysm*, in which the intimal and medial layers are disrupted and the dilatation is lined by adventitia only and sometimes by perivascular clot. Aneurysms also may be classified accordingly to their gross appearance. A *fusiform aneurysm* affects the entire circumference of a segment of the vessel, resulting in a diffusely dilated lesion. In contrast, a *saccular aneurysm* involves only a portion of the circumference, resulting in an outpouching of the vessel wall. Aortic aneurysms are also classified according to location, i.e., abdominal versus thoracic. Aneurysms of the descending thoracic aorta are usually contiguous with infradiaphragmatic aneurysms and are referred to as *thoracoabdominal aortic aneurysms*.

ETIOLOGY

The most common pathologic condition associated with aortic aneurysm is *atherosclerosis*. It is controversial whether atherosclerosis itself actually causes aortic aneurysms or whether atherosclerosis develops as a secondary event in the dilated aorta. Causality is implied by studies that have shown that many patients with aortic aneurysms have coexisting risk factors for atherosclerosis (Chap. 241), particularly cigarette smoking, as well as atherosclerosis in other blood vessels. Seventy-five percent of atherosclerotic aneurysms are located in the distal abdominal aorta, below the renal arteries.

Cystic medial necrosis is the term used to describe the degeneration of collagen and elastic fibers in the tunica media of the aorta, as well as the loss of medial cells that are replaced by multiple clefts of mucoid material. Cystic medial necrosis characteristically affects the proximal aorta, results in circumferential weakness and dilatation, and leads to development of fusiform aneurysms involving the ascending aorta and the sinuses of

Valsalva. This condition is particularly prevalent in patients with Marfan syndrome and Ehrlers-Danlos syndrome type IV (Chap. 351) but also occurs in pregnant women, in patients with hypertension, and in those with valvular heart disease. Sometimes it appears as an isolated condition in patients without any other apparent disease. Familial clusterings of aortic aneurysms occur in 20% of patients, suggesting a hereditary basis of the disease. A mutation of the gene encoding type III procollagen has been implicated. *Syphilis* (Chap. 172) is a relatively uncommon cause of aortic aneurysm. Syphilitic periaortitis and mesoaortitis damage elastic fibers, resulting in thickening and weakening of the aortic wall. Approximately 90% of syphilitic aneurysms are located in the ascending aorta or aortic arch. *Tuberculous aneurysms* (Chap. 169) typically affect the thoracic aorta and result from direct extension of infection from hilar lymph nodes or contiguous abscesses or from bacterial seeding. Loss of aortic wall elasticity results from granulomatous destruction of the medial layer. A *mycotic aneurysm* is a rare condition that develops as a result of staphylococcal, streptococcal, or salmonella infections of the aorta, usually at an atherosclerotic plaque. These aneurysms are usually saccular. Blood cultures are often positive and reveal the nature of the infecting agent.

Vasculitides associated with aortic aneurysm include Takayasu's arteritis and giant cell arteritis, which may cause aneurysms of the aortic arch and descending thoracic aorta. Spondyloarthropathies such as ankylosing spondylitis, rheumatoid arthritis, psoriatic arthritis, relapsing polychondritis, Behcet's syndrome, and Reiter's syndrome are associated with dilatation of the ascending aorta. *Traumatic aneurysms* may develop after penetrating or non-penetrating chest trauma and most commonly affect the descending thoracic aorta just beyond the site of insertion of the ligamentum arteriosum. *Congenital aortic aneurysms* may be primary or associated with anomalies such as a bicuspid aortic valve or aortic coarctation.

THORACIC AORTIC ANEURYSMS

The clinical manifestations and natural history of thoracic aortic aneurysms depend on their location. Cystic medial necrosis is the most common cause of ascending aortic aneurysms, whereas atherosclerosis is the condition most frequently associated with aneurysms of the aortic arch and descending thoracic aorta. The average growth rate of thoracic aneurysms is 0.1 to 0.4 cm per year. The risk of rupture is related to the size of the aneurysm and the presence of symptoms; it increases substantially for ascending aortic aneurysms>6 cm and descending thoracic aneurysms>7 cm. Most thoracic aortic aneurysms are asymptomatic. However, compression or erosion of adjacent tissue by aneurysms may cause symptoms such as chest pain, shortness of breath, cough, hoarseness, or dysphagia. Aneurysmal dilatation of the ascending aorta may cause congestive heart failure as a consequence of aortic regurgitation; and compression of the superior vena cava may produce congestion of the head, neck, and upper extremities.

A chest x-ray may be the first test to suggest the diagnosis of a thoracic aortic aneurysm. Findings include widening of the mediastinal shadow and displacement or compression of the trachea or left mainstem bronchus. Two-dimensional echocardiography, and particularly transesophageal echocardiography, can be used to assess the proximal ascending aorta and descending thoracic aorta. Both

contrast-enhanced computed tomography (CT) and magnetic resonance imaging (MRI) are sensitive and specific tests for assessment of aneurysms of the thoracic aorta. In asymptomatic patients whose aneurysms are too small to justify surgery, noninvasive testing with either contrast-enhanced CT or MRI should be performed at least every 6 to 12 months to monitor expansion. Contrast aortography is frequently required preoperatively to assess the length of the aneurysm and involvement of branch vessels.

Patients with thoracic aortic aneurysms, and particularly patients with Marfan syndrome who have evidence of aortic root dilatation, should receive long-term beta-blocker therapy. Additional medical therapy should be given, as necessary, to control hypertension. Operative repair with placement of a prosthetic graft is indicated in patients with symptomatic thoracic aortic aneurysms and in those in whom the aortic diameter is >6 cm. In patients with Marfan syndrome, thoracic aortic aneurysms >5 cm should be considered for surgery.

ABDOMINAL AORTIC ANEURYSMS

Abdominal aortic aneurysms occur more frequently in males than in females, and the incidence increases with age. Abdominal aortic aneurysms may affect 1 to 2% of men older than 50 years. At least 90% of all abdominal aortic aneurysms are affected by atherosclerosis, and most of these aneurysms are below the level of the renal arteries. Prognosis is related to both the size of the aneurysm and the severity of coexisting coronary artery and cerebrovascular disease. The risk of rupture increases with the size of the aneurysm. The 5-year risk of rupture for aneurysms<5 cm is 1 to 2%, whereas it is 20 to 40% for aneurysms >5 cm in diameter. The formation of mural thrombi within the aneurysm may predispose to peripheral embolization.

An abdominal aortic aneurysm commonly produces no symptoms. It is usually detected on routine examination as a palpable, pulsatile, and nontender mass, or it is an incidental finding during an abdominal x-ray or ultrasound performed for other reasons. However, as abdominal aortic aneurysms expand, they may become painful. Some patients complain of strong pulsations in the abdomen; others experience pain in the chest, lower back, or scrotum. Aneurysmal pain is usually a harbinger of rupture and represents a medical emergency. More often, acute rupture occurs without any prior warning, and this complication is always life-threatening. Rarely, there is leakage of the aneurysm with severe pain and tenderness. Acute pain and hypotension occur with rupture of the aneurysm, which requires emergency operation.

Abdominal radiography may demonstrate the calcified outline of the aneurysm. However, about 25% of aneurysms are not calcified and cannot be visualized by plain x-ray. An abdominal ultrasound can delineate the transverse and longitudinal dimensions of an abdominal aortic aneurysm and may detect mural thrombus. Abdominal ultrasound in useful for serial documentation of aneurysm size and can be used to screen patients at risk for developing aortic aneurysm, such as those with affected siblings, peripheral atherosclerosis, or peripheral artery aneurysms. CT with contrast and MRI are accurate, noninvasive tests to determine the location and size of abdominal aortic aneurysms. Contrast aortography is used commonly for the evaluation of patients with aneurysms before surgery; but the procedure carries a small risk of complications, such as bleeding, allergic reactions, and atheroembolism. This technique

is useful in documenting the length of the aneurysm, especially its upper and lower limits, and the extent of associated atherosclerotic vascular disease. However, since the presence of mural clots may reduce the luminal size, aortography may underestimate the diameter of an aneurysm.

TREATMENT

Operative repair of the aneurysm and insertion of a prosthetic graft is indicated for abdominal aortic aneurysms of any size that are expanding rapidly or are associated with symptoms. For asymptomatic aneurysms, operation is indicated if the diameter is >5 cm. Operation may be recommended in patients with aneurysm diameters of 4 to 5 cm, except for patients with exceptionally high operative risk. However, in a recent randomized trial of patients with abdominal aortic aneurysms <5.5 cm, there was no difference in the 6-year mortality rate between those followed with ultrasound surveillance and those undergoing elective aneurysm repair. Thus, serial noninvasive follow-up of smaller aneurysms (<5 cm) is an alternative to immediate surgery. Percutaneous placement of endovascular stent grafts (Fig. 247-1) for treatment of infrarenal abdominal aortic aneurysms is currently available for selected patients, and initial reports have been favorable.

In surgical candidates, careful preoperative cardiac and general medical evaluations (followed by appropriate therapy of complicating conditions) are essential. Preexisting coronary artery disease, congestive heart failure, pulmonary disease, diabetes, and advanced age add to the risk of surgery. Perioperative management should include the placement of a Swan-Ganz catheter and arterial line to monitor and optimize left ventricular filling pressure, cardiac output, and arterial pressure, especially during clamping and unclamping of the aorta, as well as during the immediate postoperative period. With careful preoperative cardiac evaluation and postoperative care the operative mortality rate approximates 1 to 2%. After acute rupture, the mortality rate of emergent operation generally exceeds 50%.

AORTIC DISSECTION

Aortic dissection is caused by a circumferential or, less frequently, transverse tear of the intima. It often occurs along the right lateral wall of the ascending aorta where the hydraulic shear stress is high. Another common site is the descending thoracic aorta just below the ligamentum arteriosum. The initiating event is either a primary intimal tear with secondary dissection into the media or a medial hemorrhage that dissects into and disrupts the intima. The pulsatile aortic flow then dissects along the elastic lamellar plates of the aorta and creates a false lumen. The dissection usually propagates distally down the descending aorta and into its major branches, but it also may propagate proximally. In some cases, a secondary distal intimal disruption occurs, resulting in the reentry of blood from the false to the true lumen.

There are at least two important pathologic and radiologic variants: intramural hematoma without an intimal flap and penetrating ulcer. The clinical picture and therapeutic management of intramural hematoma are similar to those for classic aortic dissection. By contrast, penetrating ulcers are usually localized and are not associated with extensive propagation. They are primarily found in the distal portion of the

descending thoracic aorta and are associated with extensive atherosclerotic disease. The ulcer can erode beyond the intimal border, leading to medial hematoma, and may progress to false aneurysm formation or rupture.

DeBakey and coworkers initially classified aortic dissections as type I, in which an intimal tear occurs in the ascending aorta but which involves the descending aorta as well; type II, in which the dissection is limited to the ascending aorta; and type III, in which the intimal tear is located in the descending area with distal propagation of the dissection (Fig. 247-2). Another classification (Stanford) is that of type A, in which the dissection involves the ascending aorta (proximal dissection), and type B, in which it is limited to the descending aorta (distal dissection). From a management standpoint, classification into type A or B is more practical and useful, since DeBakey types I and II are managed in a similar manner.

The factors that predispose to aortic dissection include systemic hypertension, a coexisting condition in 70% of patients, and cystic medial necrosis. Aortic dissection is the major cause of morbidity and mortality in patients with Marfan syndrome (Chap. 351) and similarly may affect patients with Ehlers-Danlos syndrome. The incidence is also increased in patients with inflammatory aortitis (i.e., Takayasu's arteritis, giant cell arteritis), congenital aortic valve anomalies (e.g., bicuspid valve), in those with coarctation of the aorta, and in otherwise normal women during the third trimester of pregnancy.

CLINICAL MANIFESTATIONS

The peak incidence is in the sixth and seventh decades. Men are more affected than women by a ratio of 2:1. The presentations of aortic dissection and its variants are the consequences of intimal tear, dissecting hematoma, occlusion of involved arteries, and compression of adjacent tissues. Acute aortic dissection presents with the sudden onset of pain (Chap. 13), which is often described as very severe and tearing and is associated with diaphoresis. The pain may be localized to the front or back of the chest, often the interscapular region, and typically migrates with propagation of the dissection. Other symptoms include syncope, dyspnea, and weakness. Physical findings may include hypertension or hypotension, loss of pulses, aortic regurgitation, pulmonary edema, and neurologic findings due to carotid artery obstruction (hemiplegia, hemianesthesia) or spinal cord ischemia (paraplegia). Bowel ischemia, hematuria, and myocardial ischemia have all been observed. These clinical manifestations reflect complications resulting from the dissection occluding the major arteries. Furthermore, clinical manifestations may result from the compression of adjacent structures (e.g., superior cervical ganglia, superior vena cava, bronchus, esophagus) by the expanding dissection causing aneursymal dilatation, and include Horner's syndrome, superior vena caval syndrome, hoarseness, dysphagia, and airway compromise. Hemopericardium and cardiac tamponade may complicate a type A lesion with retrograde dissection. Acute aortic regurgitation is an important and common (>50%) complication of proximal dissection. It is the outcome of either a circumferential tear that widens the aortic root or a disruption of the annulus by dissecting hematoma that tears a leaflet(s) or displaces it below the line of closure. Signs of aortic regurgitation include bounding pulses, a wide pulse pressure, a diastolic murmur often radiating along the right sternal border, and evidence of congestive heart failure. The clinical manifestation depends on the severity

of the regurgitation.

In dissections involving the ascending aorta, the chest x-ray often reveals a widened superior mediastinum. A pleural effusion (usually left-sided) also may be present. This effusion is typically serosanguinous and not indicative of rupture unless accompanied by hypotension and falling hematocrit. In dissections of the descending thoracic aorta, a widened mediastinum also may be observed on chest x-ray. In addition, the descending aorta may appear to be wider than the ascending portion. An electrocardiogram that shows no evidence of ischemia is helpful in distinguishing aortic dissection from myocardial infarction. Rarely, the dissection involves the right or left coronary ostium and causes acute myocardial infarction. The diagnosis of aortic dissection can be established by aortography or by the use of noninvasive techniques such as echocardiography,CT, orMRI. Aortography may be used to document the diagnosis; identify the entry point, the intimal flap, and the false and true lumina; and to establish the extent of dissection into the major arteries. Coronary angiography may be performed concomitantly in high-risk patients in the evaluation and preparation for surgery. The sensitivity of aortography is 70% for visualizing an intimal flap, 56% for the site of intimal tear, and 87% for false lumen. It is unable to recognize intramural hemorrhage. Transthoracic echocardiography can be performed simply and rapidly and has an overall sensitivity of 60 to 85%. For diagnosing proximal ascending aortic dissections, its sensitivity exceeds 80%; it is less useful for detecting dissection of the arch and descending thoracic aorta. Transesophageal echocardiography (Fig. 247-3) requires greater skill and patient cooperation but is very accurate in identifying dissections of the ascending and descending thoracic aorta, but not the arch, achieving 98% sensitivity and approximately 90% specificity. CT and MRI are both highly accurate in identifying the intimal flap and the extent of the dissection; each has a sensitivity and specificity exceeding 90%. They are useful in recognizing intramural hemorrhage and penetrating ulcers. MRI also can detect blood flow, which may be useful in characterizing antegrade versus retrograde dissection. These noninvasive tests are now becoming the diagnostic procedures of choice. Their relative utility depends on the availability and expertise in individual institutions as well as on the hemodynamic stability of the patient, with CT and MRI obviously less suitable for more unstable patients.

TREATMENT

Medical therapy should be initiated as soon as the diagnosis is considered. The patient should be admitted to an intensive care unit for monitoring hemodynamics and urine output. Unless hypotension is present, therapy should be aimed at reducing cardiac contractility and systemic arterial pressure, and thereby shear stress. For acute dissection, unless contraindicated,b-adrenergic blockers should be administered parenterally, using intravenous propranolol, metoprolol, or the short-acting esmolol to achieve a heart rate of approximately 60 beats per minute. This should be accompanied by sodium nitroprusside infusion to lower systolic blood pressure to 120 mmHg or less. Labetalol (Chap. 246), a drug with bothb- and a-adrenergic blocking properties, also has been used as a parenteral agent in the acute therapy of dissection.

The calcium channel antagonists, verapamil and diltiazem, may be used intravenously if nitroprusside or labetalol cannot be employed. Experience with calcium antagonists is limited. Direct vasodilators, such as diazoxide and hydralazine, are contraindicated

because these agents can increase hydraulic shear and may propagate dissection.

Emergent or urgent surgical correction is the preferred treatment for ascending aortic dissections (type A) and complicated type B dissections including those characterized by propagation, compromise of major aortic branches, impending rupture, or continued pain. Surgery involves excision of the intimal flap, obliteration of the false lumen, and placement of an interposition graft. A composite valve-graft conduit is used if the aortic valve is disrupted. The overall in-hospital mortality rate after surgical treatment of patients with aortic dissection is reported to be 15 to 20%. The major causes of perioperative mortality and morbidity include myocardial infarction, paraplegia, renal failure, tamponade, hemorrhage, and sepsis. Recent reports of the use of endoluminal stent grafts in selected patients with type B dissection have been encouraging. Other transcatheter techniques, such as fenestration of the intimal flaps and stenting of narrowed branch vessels to increase flow to compromised organs, are also under investigation. For uncomplicated and stable distal dissection (type B), medical therapy is the preferred treatment. The in-hospital mortality rate of medically treated patients with type B dissection is 15 to 20%. Long-term therapy for patients with aortic dissection (with or without surgery) consists of the control of hypertension and reduction of cardiac contractility with the use of beta blockers plus other antihypertensive agents such as angiotensin-converting enzyme inhibitor or calcium antagonist. Patients with chronic type B dissection should be followed on an outpatient basis every 6 to 12 months by contrast-enhancedCT orMRI to detect propagation. Patients with Marfan syndrome are at high risk for postdissection complications. The long-term prognosis for patients with treated dissections is generally good with careful follow-up; the 10-year survival rate is approximately 60%.

AORTIC OCCLUSION

CHRONIC ATHEROSCLEROTIC OCCLUSIVE DISEASE

Atherosclerosis may affect the thoracic and abdominal aorta, but occlusive aortic disease caused by atherosclerosis usually is confined to the distal abdominal aorta below the renal arteries. Frequently the disease extends to the iliac arteries (Chap. 248). Claudication characteristically involves the lower back, buttocks, and thighs and may be associated with impotence in males (Leriche syndrome). The severity of the symptoms depends on the adequacy of collaterals. With sufficient collateral blood flow, a complete occlusion of the abdominal aorta may occur without the development of ischemic symptoms. The physical findings include absence of femoral and other distal pulses bilaterally and the detection of an audible bruit over the abdomen (usually at or below the umbilicus) and the common femoral arteries. Atrophic skin, loss of hair, and coolness of the lower extremities are usually observed. In advanced ischemia, rubor on dependency and pallor on elevation can be seen.

The diagnosis is usually established by the physical examination and noninvasive testing, including leg pressure measurements, Doppler velocity analysis, and pulse volume recordings. The anatomy may be defined by abdominal aortography before revascularization. Operative treatment is indicated in patients with debilitating symptoms and/or with the development of leg ischemia.

ACUTE OCCLUSION

Acute occlusion in the distal abdominal aorta represents a medical emergency because it threatens the viability of the lower extremities. It usually results from an occlusive embolus that almost always originates from the heart. Rarely, acute occlusion may occur as the result of in situ thrombosis in a preexisting severely narrowed segment of the aorta or plaque rupture and hemorrhage into such an area.

The clinical picture is one of acute ischemia of the lower extremities. Severe rest pain, coolness, and pallor of the lower extremities and the absence of distal pulses bilaterally are the usual manifestations. Diagnosis should be established rapidly by aortography. Emergency thrombectomy or revascularization is indicated.

AORTITIS

Aortitis frequently affects the ascending aorta and may result in aneurysmal dilatation and aortic regurgitation; it occasionally obstructs branch vessels of the aorta.

SYPHILITIC AORTITIS

This late manifestation of luetic infection (Chap. 172) usually affects the proximal ascending aorta, particularly the aortic root, resulting in aortic dilatation and aneurysm formation. Syphilitic aortitis may occasionally involve the aortic arch or the descending aorta. The aneurysms may be saccular or fusiform and are usually asymptomatic, but compression of and erosion into adjacent structures may result in symptoms; rupture also may occur.

The initial lesion is an obliterative endarteritis of the vasa vasorum, especially in the adventitia. This is an inflammatory response to the invasion of the adventitia by the spirochetes. Destruction of the aortic media occurs as the spirochetes spread into this layer, usually via the lymphatics accompanying the vasa vasorum. Destruction of collagen and elastic tissues leads to dilation of the aorta, scar formation, and calcification. These changes account for the characteristic radiographic appearance of a calcified ascending aortic aneurysm.

The disease typically presents as an incidental radiographic finding 15 to 30 years after initial infection. Symptoms may result from aortic regurgitation, narrowing of coronary ostia due to syphilitic aortitis, compression of adjacent structures (e.g., esophagus), or rupture. Diagnosis is established by a positive serologic test, i.e., rapid plasmin reagin (RPR) or fluorescent treponemal antibody. Treatment includes penicillin and surgical excision and repair.

RHEUMATIC AORTITIS

Rheumatoid arthritis (Chap. 312), ankylosing spondylitis (Chap. 315), psoriatic arthritis (Chap. 324), Reiter's syndrome (Chap. 315), Behcet's syndrome (Chap. 316), relapsing polychondritis, and inflammatory bowel disorders may all be associated with aortitis involving the ascending aorta. The inflammatory lesions usually involve the ascending aorta and may extend to the sinuses of Valsalva, the mitral valve leaflets, and adjacent

myocardium. The clinical manifestations are aneurysm, aortic regurgitation, and involvement of the cardiac conduction system.

TAKAYASU'S ARTERITIS

Inflammatory diseases of the aortic arch resulting in obstruction of the aorta and its major arteries characterize this major group of diseases. Takayasu's arteritis is also termed *pulseless disease* because of the frequent occlusion of the large arteries originating from the aorta. It also may involve the descending thoracic and abdominal aorta and occlude large branches such as the renal arteries. Aortic aneurysms may also occur. The pathology is a panarteritis, characterized by mononuclear cells and occasionally giant cells, with marked intimal hyperplasia, medial and adventitial thickening, and, in chronic form, fibrotic occlusion. The disease is most prevalent in young females of Asian descent. During the acute stage, fever, malaise, weight loss, and other systemic symptoms may be evident. An elevation of the erythrocyte sedimentation rate is common. The chronic stages of the disease present with symptoms related to large artery occlusion, such as upper extremity claudication, cerebral ischemia, and syncope. The chronic disease is intermittently active. Since the process is progressive and there is no definitive therapy, the prognosis is usually poor. Glucocorticoids and immunosuppressive agents have been reported to be effective in some patients during the acute phase. Occasionally, anticoagulation prevents thrombosis and complete occlusion of a large artery. Surgical bypass of a critically stenotic artery may be necessary.

GIANT CELL ARTERITIS (See also Chap. 317)

This vasculitis occurs in older individuals and affects women more often than men. Primarily large and medium-sized arteries are affected. The pathology is that of focal granulomatous lesions involving the entire arterial wall. It may be associated with polymyalgia rheumatica. Obstruction of medium-sized arteries (e.g., temporal and ophthalmic arteries) and of major branches of the aorta and the development of aortitis and aortic regurgitation are some of the complications of the disease. High-dose glucocorticoid therapy may be effective when given early.

(Bibliography omitted in Palm version)

Back to Table of Contents

248. VASCULAR DISEASES OF THE EXTREMITIES - *Mark A. Creager*, *Victor J. Dzau*

ARTERIAL DISORDERS

PERIPHERAL ARTERIAL DISEASE

Atherosclerosis (arteriosclerosis obliterans) is the leading cause of occlusive arterial disease of the extremities in patients over 40 years old; the highest incidence occurs in the sixth and seventh decades of life. As in patients with atherosclerosis of the coronary and cerebral vasculature, there is an increased prevalence of peripheral atherosclerotic disease in individuals with diabetes mellitus, hypercholesterolemia, hypertension, or hyperhomocysteinemia and in cigarette smokers.

Pathology (See alsoChap. 241) Segmental lesions causing stenosis or occlusion are usually localized in large and medium-sized vessels. The pathology of the lesions includes atherosclerotic plaques with calcium deposition, thinning of the media, patchy destruction of muscle and elastic fibers, fragmentation of the internal elastic lamina, and thrombi composed of platelets and fibrin. The primary sites of involvement are the abdominal aorta and iliac arteries (30% of symptomatic patients), the femoral and popliteal arteries (80 to 90% of patients), and the more distal vessels, including the tibial and peroneal arteries (40 to 50% of patients). Atherosclerotic lesions occur preferentially at arterial branch points, sites of increased turbulence, altered shear stress, and intimal injury. Involvement of the distal vasculature is most common in elderly individuals and patients with diabetes mellitus.

Clinical Evaluation The most common *symptom* is intermittent claudication, which is defined as a pain, ache, cramp, numbness, or a sense of fatigue in the muscles; it occurs during exercise and is relieved by rest. The site of claudication is distal to the location of the occlusive lesion. For example, buttock, hip, and thigh discomfort occurs in patients with aortoiliac disease (Leriche syndrome), whereas calf claudication develops in patients with femoral-popliteal disease. Symptoms are far more common in the lower than in the upper extremities because of the higher incidence of obstructive lesions in the former region. In patients with severe arterial occlusive disease, critical limb ischemia may develop. Patients will complain of rest pain or a feeling of cold or numbness in the foot and toes. Frequently, these symptoms occur at night when the legs are horizontal and improve when the legs are in a dependent position. With severe ischemia, rest pain may be persistent.

Important *physical findings* of peripheral arterial disease include decreased or absent pulses distal to the obstruction, the presence of bruits over the narrowed artery, and muscle atrophy. With more severe disease, hair loss, thickened nails, smooth and shiny skin, reduced skin temperature, and pallor or cyanosis are frequent physical signs. In addition, ulcers or gangrene may occur. Elevation of the legs and repeated flexing of the calf muscles produce pallor of the soles of the feet, whereas rubor, secondary to reactive hyperemia, may develop when the legs are dependent. The time required for rubor to develop or for the veins in the foot to fill when the patient's legs are transferred from an elevated to a dependent position is related to the severity of the ischemia and the presence of collateral vessels. Patients with severe ischemia may develop

peripheral edema because they keep their legs in a dependent position much of the time. Ischemic neuritis can result in numbness and hyporeflexia.

Noninvasive Testing The history and physical examination are usually sufficient to establish the diagnosis of peripheral arterial disease. An objective assessment of the severity of disease is obtained by noninvasive techniques. These include digital pulse volume recordings, Doppler flow velocity waveform analysis, duplex ultrasonography (which combines B-mode imaging and pulse-wave Doppler examination), segmental pressure measurements, transcutaneous oximetry, stress testing (usually using a treadmill), and tests of reactive hyperemia. In the presence of significant peripheral arterial disease, the volume displacement in the leg is decreased with each pulse, and the Doppler velocity contour becomes progressively flatter. Duplex ultrasonography is often useful in detecting stenotic lesions in native arteries and bypass grafts.

Arterial pressure can be recorded noninvasively along the legs by serial placement of sphygmomanometric cuffs and use of a Doppler device to auscultate or record blood flow. Normally, systolic blood pressure in the legs and arms is similar. Indeed, ankle pressure may be slightly higher than arm pressure due to pulse-wave reflection. In the presence of hemodynamically significant stenoses, the systolic blood pressure in the leg is decreased. Thus, if one were to obtain a ratio of the ankle and brachial artery pressures, it would be ≥ 1.0 in normal individuals and <1.0 in patients with peripheral arterial disease. A ratio of <0.5 is consistent with severe ischemia.

Treadmill testing allows the physician to assess functional limitations objectively. Decline of the ankle-brachial systolic pressure ratio immediately after exercise may provide further support for the diagnosis of peripheral arterial disease in patients with equivocal symptoms and findings on examination. Exercise testing also allows simultaneous evaluation for the presence of coronary artery disease.

Contrast angiography should not be used for routine diagnostic testing but is performed prior to potential revascularization. It is useful in defining the anatomy to assist operative planning and is also indicated if nonsurgical interventions are being considered, such as percutaneous transluminal angioplasty (PTA) or thrombolysis. Recent studies have suggested that magnetic resonance angiography has diagnostic accuracy comparable to that of contrast angiography.

Prognosis The natural history of patients with peripheral arterial disease is influenced primarily by the extent of coexisting coronary artery and cerebral vascular disease. Studies using coronary angiography have estimated that approximately one-half of patients with symptomatic peripheral arterial disease also have significant coronary artery disease. Life-table analysis has indicated that patients with claudication have a 70% 5-year and a 50% 10-year survival rate. Most deaths are either sudden or secondary to myocardial infarction. The likelihood of symptomatic progression of peripheral arterial disease appears less than the chance of succumbing to coronary artery disease. Approximately 75% of nondiabetic patients who present with mild to moderate claudication remain symptomatically stable or improve. Deterioration is likely to occur in the remainder, with approximately 5% of the group ultimately undergoing amputation. The prognosis is worse in patients who continue to smoke cigarettes or who have diabetes mellitus.

TREATMENT

Therapeutic options include supportive measures, pharmacologic treatment, nonoperative interventions, and surgery. Supportive measures include meticulous care of the feet, which should be kept clean and protected against excessive drying with moisturizing creams. Well-fitting and protective shoes are advised to reduce trauma. Sandals and shoes made of synthetic materials that do not "breathe" should be avoided. Elastic support hose should be avoided, as they reduce blood flow to the skin. In patients with ischemia at rest, shock blocks under the head of the bed together with a canopy over the feet may improve perfusion pressure and ameliorate some of the rest pain.

Treatment of associated factors that contribute to the development of atherosclerosis should be initiated. The importance of discontinuing cigarette smoking cannot be overemphasized. The physician must assume a major role in this life-style modification. It is important to control blood pressure in hypertensive patients but to avoid hypotensive levels. Treatment of hypercholesterolemia is advocated, although reduction in cholesterol levels has not been shown unequivocally to reverse peripheral atherosclerotic lesions. However, it has been shown to prevent or to slow progression of the disease and to improve survival in patients with coronary atherosclerosis. Patients with claudication should also be encouraged to exercise regularly and at progressively more strenuous levels. Supervised exercise training programs may improve muscle efficiency and prolong walking distance. Patients also should be advised to walk for 30 to 45 min daily, stopping at the onset of claudication and resting until the symptoms resolve before resuming ambulation.

Pharmacologic Management This form of treatment of patients with peripheral arterial disease has not been as successful as the medical treatment of coronary artery disease (Chap. 244). In particular, vasodilators as a class have not proved to be beneficial. During exercise, peripheral vasodilation occurs distal to sites of significant arterial stenoses. As a result, perfusion pressure falls, often to levels less than that generated in the interstitial tissue by the exercising muscle. Drugs such asa-adrenergic blocking agents, calcium channel antagonists, papaverine, and other vasodilators have not been shown to be effective in patients with peripheral arterial disease. Pentoxifylline, a substituted xanthine derivative, has been reported to decrease blood viscosity and to increase red cell flexibility, thereby increasing blood flow to the microcirculation and enhancing tissue oxygenation. Several placebo-controlled studies have reported that pentoxifylline increased the duration of exercise in patients with claudication, but its efficacy has not been confirmed in all clinical trials. Cilostazol, a phosphodiesterase inhibitor with vasodilator and antiplatelet properties, has been reported to increase claudication distance and recently received an indication for treatment of peripheral arterial disease by the U.S. Food and Drug Administration. Other drugs are being studied that potentially may improve claudication, such as L-arginine, which is the precursor of the endothelium-dependent vasodilator, nitric oxide, and vasodilator prostaglandins. Several studies have suggested that long-term parenteral administration of vasodilator prostaglandins decreases pain and facilitates healing of ulcers in patients with severe limb ischemia. Clinical trials with angiogenic growth factors such as vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF) are

proceeding. A preliminary report suggested that intramuscular gene transfer of DNA encoding VEGF may promote collateral blood vessel growth in patients with critical limb ischemia.

Platelet inhibitors, particularly aspirin, reduce the risk of adverse cardiovascular events in patients with peripheral atherosclerosis. Clopidogril, a drug that inhibits platelet aggregation via its effect on ADP-dependent platelet-fibrinogen binding, appears to be more effective than aspirin in reducing cardiovascular morbidity and mortality in patients with peripheral arterial disease. The anticoagulants heparin and warfarin have not been shown to be effective in patients with chronic peripheral arterial disease but may be useful in acute arterial obstruction secondary to thrombosis or systemic embolism. Similarly, thrombolytic intervention using drugs such as streptokinase, urokinase, or recombinant tissue plasminogen activator (tPA) may have a role in the treatment of acute thrombotic arterial occlusion but is not effective in patients with chronic arterial occlusion secondary to atherosclerosis.

Revascularization Revascularization procedures, including nonoperative as well as operative interventions, are usually reserved for patients with progressive, severe, or disabling symptoms and ischemia at rest, as well as for individuals who must be symptom-free because of their occupation. Angiography should be performed mainly in patients who are being considered for a revascularization procedure. Nonoperative interventions includePTA, stent placement, and atherectomy (Chap. 245). PTA of the iliac artery is associated with a higher success rate than PTA of the femoral and popliteal arteries. Approximately 90 to 95% of iliac PTAs are initially successful, and the 3-year patency rate is in excess of 75%. Patency rates may be higher if a stent is placed in the iliac artery. The initial success rate for femoral-popliteal PTA is approximately 80%, with a 60% 3-year patency rate. Patency rates are influenced by the severity of pretreatment stenoses; the prognosis of total occlusive lesions is worse than that of nonocclusive stenotic lesions.

Several operative procedures are available for treating patients with aortoiliac and femoral-popliteal artery disease. The preferred operative procedure depends on the location and extent of the obstruction(s) and general medical condition of the patient. Operative procedures for aortoiliac disease include aortobifemoral bypass, axillofemoral bypass, femoral-femoral bypass, and aortoiliac endarterectomy. The most frequently used procedure is the aortobifemoral bypass using knitted Dacron grafts. Immediate graft patency approaches 99%, and 5- and 10-year graft patency in survivors is in excess of 90 and 80%, respectively. Operative complications include myocardial infarction and stroke, infection of the graft, peripheral embolization, and sexual dysfunction from interruption of autonomic nerves in the pelvis. Operative mortality ranges from 1 to 3%, mostly due to ischemic heart disease.

Operative therapy for femoral-popliteal artery disease includes in situ and reverse autogenous saphenous vein bypass grafts, placement of polytetrafluoroethylene (PTFE) or other synthetic grafts, and thromboendarterectomy. Operative mortality ranges from 1 to 3%. The long-term patency rate depends on the type of graft used, the location of the distal anastomosis, and the patency of runoff vessels beyond the anastomosis. Patency rates of femoral-popliteal saphenous vein bypass grafts at 1 year approach 90% and at 5 years, 70 to 80%. Five-year patency rates of infrapopliteal saphenous vein bypass

grafts are 60 to 70%. In contrast, 5-year patency rates of infrapopliteal PTFE grafts are less than 30%. Lumbar sympathectomy alone or as an adjunct to aortofemoral reconstruction has fallen into disfavor.

Preoperative cardiac risk assessment may identify individuals especially likely to experience an adverse cardiac event during the perioperative period. Patients with angina, prior myocardial infarction, ventricular ectopy, heart failure, or diabetes are among those at increased risk. Noninvasive tests, such as treadmill testing (if feasible), dipyridamole thallium or sestamibi scintigraphy, dobutamine echocardiography, and ambulatory ischemia monitoring permit further stratification of patient risk (Chap. 245). Patients with abnormal test results require close supervision and adjunctive management with antianginal medications. It is not known whether coronary angiography and coronary arterial revascularization reduce overall perioperative mortality in high-risk patients undergoing peripheral vascular surgery, but cardiac catheterization should be considered in patients suspected of having left main or three-vessel coronary artery disease.

FIBROMUSCULAR DYSPLASIA

This is a hyperplastic disorder affecting medium-sized and small arteries. It occurs predominantly in females and usually involves renal and carotid arteries but can affect extremity vessels such as the iliac and subclavian arteries. The histologic classification includes intimal, medial, and periadventitial dysplasia. Medial dysplasia is the most common type and is characterized by hyperplasia of the media with or without fibrosis of the elastic membrane. It is identified angiographically by a "string of beads" appearance caused by thickened fibromuscular ridges contiguous with thin, less involved portions of the arterial wall. When limb vessels are involved, clinical manifestations are similar to those for atherosclerosis, including claudication and rest pain.PTA and surgical reconstruction have been beneficial in patients with debilitating symptoms or threatened limbs.

THROMBOANGIITIS OBLITERANS

Thromboangiitis obliterans (Buerger's disease) is an inflammatory occlusive vascular disorder involving small and medium-sized arteries and veins in the distal upper and lower extremities. Cerebral, visceral, and coronary vessels may also be affected. This disorder develops most frequently in men under age 40. The prevalence is higher in Asians and individuals of eastern European descent. While the cause of thromboangiitis obliterans is not known, there is a definite relationship to cigarette smoking in patients with this disorder.

In the initial stages of thromboangiitis obliterans, polymorphonuclear leukocytes infiltrate the walls of the small and medium-sized arteries and veins. The internal elastic lamina is preserved, and thrombus may develop in the vascular lumen. As the disease progresses, mononuclear cells, fibroblasts, and giant cells replace the neutrophils. Later stages are characterized by perivascular fibrosis and recanalization.

The clinical features of thromboangiitis obliterans often include a triad of claudication of the affected extremity, Raynaud's phenomenon (p. 1438), and migratory superficial vein

thrombophlebitis. Claudication is usually confined to the calves and feet or the forearms and hands, because this disorder primarily affects distal vessels. In the presence of severe digital ischemia, trophic nail changes, painful ulcerations, and gangrene may develop at the tips of the fingers. The physical examination shows normal brachial and popliteal pulses but reduced or absent radial, ulnar, and/or tibial pulses. Arteriography is helpful in making the diagnosis. Smooth, tapering segmental lesions in the distal vessels are characteristic, as are collateral vessels at sites of vascular occlusion. Proximal atherosclerotic disease is usually absent. The diagnosis can be confirmed by excisional biopsy and pathologic examination of an involved vessel.

There is no specific treatment except abstention from tobacco. The prognosis is worse in individuals who continue to smoke, but results are discouraging even in those who do stop smoking. Arterial bypass of the larger vessels may be used in selected instances, as well as local debridement, depending on the symptoms and severity of ischemia. Antibiotics may be useful; anticoagulants and glucocorticoids are not helpful. If these measures fail, amputation may be required.

VASCULITIS

Other vasculitides may affect the arteries supplying the upper and lower extremities.**Takayasu's arteritis and giant cell (temporal) arteritis are discussed in Chap. 317.*

ACUTE ARTERIAL OCCLUSION

This results in the sudden cessation of blood flow to an extremity. The severity of ischemia and the viability of the extremity depend on the location and extent of the occlusion and the presence and subsequent development of collateral blood vessels. There are two principal causes of acute arterial occlusion: embolism and thrombus in situ.

The most common sources of arterial emboli are the heart, aorta, and large arteries. Cardiac disorders that cause thromboembolism include atrial fibrillation, both chronic and paroxysmal; acute myocardial infarction; ventricular aneurysm; cardiomyopathy; infectious and marantic endocarditis; prosthetic heart valves; and atrial myxoma. Emboli to the distal vessels may also originate from proximal sites of atherosclerosis and aneurysms of the aorta and large vessels. Less frequently, an arterial occlusion results paradoxically from a venous thrombus that has entered the systemic circulation via a patent foramen ovale or other septal defect. Arterial emboli tend to lodge at vessel bifurcations because the vessel caliber decreases at these sites; in the lower extremities, emboli lodge most frequently in the femoral artery, followed by the iliac artery, aorta, and popliteal and tibioperoneal arteries.

Acute arterial thrombosis in situ occurs most frequently in atherosclerotic vessels at the site of a stenosis or aneurysm and in arterial bypass grafts. Trauma to an artery may also result in the formation of an acute arterial thrombus. Arterial occlusion may complicate arterial punctures and placement of catheters. Less frequent causes include the thoracic outlet compression syndrome, which causes subclavian artery occlusion, and entrapment of the popliteal artery by abnormal placement of the medial head of the

gastrocnemius muscle. Polycythemia and hypercoagulable disorders (Chaps. 110 and118) are also associated with acute arterial thrombosis.

Clinical Features The symptoms of an acute arterial occlusion depend on the location, duration, and severity of the obstruction. Often, severe pain, paresthesia, numbness, and coldness develop in the involved extremity within 1 h. Paralysis may occur with severe and persistent ischemia. Physical findings include loss of pulses distal to the occlusion, cyanosis or pallor, mottling, decreased skin temperature, muscle stiffening, loss of sensation, weakness, and/or absent deep tendon reflexes. If acute arterial occlusion occurs in the presence of an adequate collateral circulation, as is often the case in acute graft occlusion, the symptoms and findings may be less impressive. In this situation, the patient complains about an abrupt decrease in the distance walked before claudication occurs or of modest pain and paresthesia. Pallor and coolness are evident, but sensory and motor functions are generally preserved. The diagnosis of acute arterial occlusion is usually apparent from the clinical presentation. Arteriography is useful for confirming the diagnosis and demonstrating the location and extent of occlusion.

TREATMENT

Once the diagnosis is made, the patient should be anticoagulated with intravenous heparin to prevent propagation of the clot. In cases of severe ischemia of recent onset, and particularly when limb viability is jeopardized, immediate intervention to ensure reperfusion is indicated. Surgical thromboembolectomy or arterial bypass procedures are used to restore blood flow to the ischemic extremity promptly, particularly when a large proximal vessel is occluded.

Intraarterial thrombolytic therapy is effective when acute arterial occlusion is caused by a thrombus in an atherosclerotic vessel or arterial bypass graft. Thrombolytic therapy may also be indicated when the patient's overall condition contraindicates surgical intervention or when smaller distal vessels are occluded, thus preventing surgical access. One approach for administering intraarterial urokinase is to give 240,000 IU/h for 4 h, followed by 120,000 IU/h for a maximum of 48 h. Intraarterial recombinanttPA may be administered at infusion rates of 1 mg/h or 0.05 mg/kg per hour. Meticulous observation for hemorrhagic complications is required during intraarterial thrombolytic therapy.

If the limb is not in jeopardy, a more conservative approach that includes observation and administration of anticoagulants may be taken. Anticoagulation prevents recurrent embolism and reduces the likelihood of thrombus propagation. It can be initiated with intravenous heparin and followed by oral warfarin. Recommended dosages are the same as those used for deep vein thrombosis (see below). Emboli resulting from infectious endocarditis, the presence of prosthetic heart valves, or atrial myxoma often require surgical intervention to remove the cause.

ATHEROEMBOLISM

Atheroembolism constitutes a subset of acute arterial occlusion. In this condition, multiple small deposits of fibrin, platelet, and cholesterol debris embolize from proximal atherosclerotic lesions or aneurysmal sites. Atheroembolism may occur after

intraarterial procedures. Since the emboli tend to lodge in the small vessels of the muscle and skin and may not occlude the large vessels, distal pulses usually remain palpable. Patients complain of acute pain and tenderness at the site of embolization. Digital vascular occlusion may result in ischemia and the "blue toe" syndrome; digital necrosis and gangrene may develop. Localized areas of tenderness, pallor, and livedo reticularis (see below) occur at sites of emboli. Skin or muscle biopsy may demonstrate cholesterol crystals.

Ischemia resulting from atheroemboli is notoriously difficult to treat. Usually neither surgical revascularization procedures nor thrombolytic therapy is helpful because of the multiplicity, composition, and distal location of the emboli. Some evidence suggests that platelet inhibitors prevent atheroembolism. Surgical intervention to remove or bypass the atherosclerotic vessel or aneurysm that causes the recurrent atheroemboli may be necessary.

THORACIC OUTLET COMPRESSION SYNDROME

This is a symptom complex resulting from compression of the neurovascular bundle (artery, vein, or nerves) at the thoracic outlet as it courses through the neck and shoulder. Cervical ribs, abnormalities of the scalenus anticus muscle, proximity of the clavicle to the first rib, or abnormal insertion of the pectoralis minor muscle may compress the subclavian artery and brachial plexus as these structures pass from the thorax to the arm. Patients may develop shoulder and arm pain, weakness, paresthesia, claudication, Raynaud's phenomenon, and even ischemic tissue loss and gangrene. Examination is often normal unless provocative maneuvers are performed. Occasionally, distal pulses are decreased or absent and digital cyanosis and ischemia may be evident. Tenderness may be present in the supraclavicular fossa. Abducting the affected arm by 90° and externally rotating the shoulder may precipitate symptoms. Several additional maneuvers are used to confirm the diagnosis of vascular compression and to suggest the location of the abnormality. These include the scalene maneuver (extension of the neck and rotation of the head to the side of the symptoms), the costoclavicular maneuver (posterior rotation of shoulders), and the hyperabduction maneuver (raising the arm 180°), which may cause subclavian bruits and loss of pulses in the arm. A chest x-ray will indicate the presence of cervical ribs. The electromyogram will be abnormal if the brachial plexus is involved.

TREATMENT

Most patients can be managed conservatively. They should be advised to avoid the positions that cause symptoms. Many patients benefit from shoulder girdle exercises. Surgical procedures such as removal of the first rib or resection of the scalenus anticus muscle are necessary occasionally for relief of symptoms or treatment of ischemia.

ARTERIOVENOUS FISTULA

Abnormal communications between an artery and a vein, bypassing the capillary bed, may be congenital or acquired. Congenital arteriovenous fistulas are the result of persistent embryonic vessels that fail to differentiate into arteries and veins; they may be associated with birthmarks, can be located in almost any organ of the body, and

frequently occur in the extremities. Acquired arteriovenous fistulas are either created to provide vascular access for hemodialysis or occur as a result of a penetrating injury such as a gunshot or knife wound or as complications of arterial catheterization or surgical dissection. An infrequent cause of arteriovenous fistula is rupture of an arterial aneurysm into a vein.

The clinical features depend on the location and size of the fistula. Frequently, a pulsatile mass is palpable, and a thrill and bruit lasting throughout systole and diastole are present over the fistula. With long-standing fistulas, clinical manifestations of chronic venous insufficiency, including peripheral edema, large, tortuous varicose veins, and stasis pigmentation become apparent because of the high venous pressure. Evidence of ischemia may occur in the distal portion of the extremity. Skin temperature is higher over the arteriovenous fistula. Large arteriovenous fistulas may result in an increased cardiac output with consequent cardiomegaly and high-output heart failure (Chap. 232).

Diagnosis The diagnosis is often evident from the physical examination. Compression of a large arteriovenous fistula may cause reflex slowing of the heart rate (Nicoladoni-Branham sign). Arteriography can confirm the diagnosis and is useful in demonstrating the site and size of the arteriovenous fistula.

TREATMENT

Management of arteriovenous fistulas may involve surgery, radiotherapy, or embolization. Congenital arteriovenous fistulas are often difficult to treat because the communications may be numerous and extensive, and new ones frequently develop after ligation of the most obvious ones. Many of these lesions are best treated conservatively using elastic support hose to reduce the consequences of venous hypertension. Occasionally, embolization with autologous material, such as fat or muscle, or with hemostatic agents, such as gelatin sponges or silicon spheres, is used to obliterate the fistula. Acquired arteriovenous fistulas are usually amenable to surgical treatment that involves division or excision of the fistula. Occasionally, autogenous or synthetic grafting is necessary to reestablish continuity of the artery and vein.

RAYNAUD'S PHENOMENON

Raynaud's phenomenon is characterized by episodic digital ischemia, manifested clinically by the sequential development of digital blanching, cyanosis, and rubor of the fingers or toes following cold exposure and subsequent rewarming. Emotional stress may also precipitate Raynaud's phenomenon. The color changes are usually well demarcated and are confined to the fingers or toes. Typically, one or more digits will appear white when the patient is exposed to a cold environment or touches a cold object. The blanching, or pallor, represents the ischemic phase of the phenomenon and results from vasospasm of digital arteries. During the ischemic phase, capillaries and venules dilate, and cyanosis results from the deoxygenated blood that is present in these vessels. A sensation of cold or numbness or paresthesia of the digits often accompanies the phases of pallor and cyanosis.

With rewarming, the digital vasospasm resolves, and blood flow into the dilated arterioles and capillaries increases dramatically. This "reactive hyperemia" imparts a

bright red color to the digits. In addition to rubor and warmth, patients often experience a throbbing, painful sensation during the hyperemic phase. Although the triphasic color response is typical of Raynaud's phenomenon, some patients may develop only pallor and cyanosis; others may experience only cyanosis.

Pathophysiology Raynaud originally proposed that cold-induced episodic digital ischemia was secondary to exaggerated reflex sympathetic vasoconstriction. This theory is supported by the fact that a-adrenergic blocking drugs as well as sympathectomy decrease the frequency and severity of Raynaud's phenomenon in some patients. An alternative hypothesis is that the digital vascular responsiveness to cold or to normal sympathetic stimuli is enhanced. It is also possible that normal reflex sympathetic vasoconstriction is superimposed on local digital vascular disease or that there is enhanced adrenergic neuroeffector activity.

Raynaud's phenomenon is broadly separated into two categories: the idiopathic variety, termed *Raynaud's disease*, and the secondary variety, which is associated with other disease states or known causes of vasospasm (Table 248-1).

Raynaud's Disease This appellation is applied when the secondary causes of Raynaud's phenomenon have been excluded. Over 50% of patients with Raynaud's phenomenon have Raynaud's disease. Women are affected about five times more often than men, and the age of presentation is usually between 20 and 40 years. The fingers are involved more frequently than the toes. Initial episodes may involve only one or two fingertips, but subsequent attacks may involve the entire finger and may include all the fingers. The toes are affected in 40% of patients. Although vasospasm of the toes usually occurs in patients with symptoms in the fingers, it may happen alone. Rarely, the earlobes and the tip of the nose are involved. Raynaud's phenomenon occurs frequently in patients who also have migraine headaches or variant angina. These associations suggest that there may be a common predisposing cause for the vasospasm.

Results of physical examination often are entirely normal; the radial, ulnar, and pedal pulses are normal. The fingers and toes may be cool between attacks and may perspire excessively. Thickening and tightening of the digital subcutaneous tissue (*sclerodactyly*) develop in 10% of patients. Angiography of the digits for diagnostic purposes is not indicated.

In general, patients with Raynaud's disease appear to have the milder forms of Raynaud's phenomenon. Fewer than 1% of these patients lose a part of a digit. After the diagnosis is made, the disease improves spontaneously in approximately 15% of patients and progresses in about 30%.

Secondary Causes of Raynaud's Phenomenon Raynaud's phenomenon occurs in 80 to 90% of patients with systemic sclerosis (scleroderma) and is the presenting symptom in 30% (Chap. 313). It may be the only symptom of scleroderma for many years. Abnormalities of the digital vessels may contribute to the development of Raynaud's phenomenon in this disorder. Ischemic fingertip ulcers may develop and progress to gangrene and autoamputation. About 20% of patients with systemic lupus erythematosus (SLE) have Raynaud's phenomenon (Chap. 311). Occasionally, persistent digital ischemia develops and may result in ulcers or gangrene. In most

severe cases, the small vessels are occluded by a proliferative endarteritis. Raynaud's phenomenon occurs in about 30% of patients with dermatomyositis or polymyositis (Chap. 382). It frequently develops in patients with rheumatoid arthritis and may be related to the intimal proliferation that occurs in the digital arteries.

Atherosclerosis of the extremities is a frequent cause of Raynaud's phenomenon in men over age 50. Thromboangiitis obliterans is an uncommon cause of Raynaud's phenomenon but should be considered in young men, particularly in those who are cigarette smokers. The development of cold-induced pallor in these disorders may be confined to one or two digits of the involved extremity. Occasionally, Raynaud's phenomenon may follow acute occlusion of large and medium-sized arteries by a thrombus or embolus. Embolization of atheroembolic debris may cause digital ischemia. The latter situation often involves one or two digits and should not be confused with Raynaud's phenomenon. In patients with the thoracic outlet syndrome, Raynaud's phenomenon may result from diminished intravascular pressure, stimulation of sympathetic fibers in the brachial plexus, or a combination of both. Raynaud's phenomenon occurs in patients with primary pulmonary hypertension (Chap. 260); this is more than coincidental and may reflect a neurohumoral abnormality that affects both the pulmonary and digital circulations.

A variety of blood dyscrasias may be associated with Raynaud's phenomenon. Cold-induced precipitation of plasma proteins, hyperviscosity, and aggregation of red cells and platelets may occur in patients with cold agglutinins, cryoglobulinemia, or cryofibrinogenemia. Hyperviscosity syndromes that accompany myeloproliferative disorders and Waldenstrom's macroglobulinemia should also be considered in the initial evaluation of patients with Raynaud's phenomenon.

Raynaud's phenomenon occurs often in patients whose vocations require the use of vibrating hand tools, such as chain saws or jackhammers. The frequency of Raynaud's phenomenon also seems to be increased in pianists and typists. Electric shock injury to the hands or frostbite may lead to the later development of Raynaud's phenomenon.

Several drugs have been causally implicated in Raynaud's phenomenon. These include ergot preparations, methysergide, b-adrenergic receptor antagonists, and the chemotherapeutic agents bleomycin, vinblastine, and cisplatin.

TREATMENT

Most patients with Raynaud's phenomenon experience only mild and infrequent episodes. These patients need reassurance and should be instructed to dress warmly and avoid unnecessary cold exposure. In addition to gloves and mittens, patients should protect the trunk, head, and feet with warm clothing to prevent cold-induced reflex vasoconstriction. Tobacco use is contraindicated.

Drug treatment should be reserved for the severe cases. The calcium channel antagonists, especially nifedipine and diltiazem, decrease the frequency and severity of Raynaud's phenomenon. Adrenergic blocking agents, such as reserpine, have been shown to increase nutritional blood flow to the fingers. Some, but not all, patients achieve satisfactory results with long-term reserpine therapy. Moreover, systemic use of

this drug is limited by side effects of hypotension, nasal stuffiness, lethargy, and depression. The postsynaptica$_1$-adrenergic antagonist prazosin has been used with favorable responses. Doxazosin and terazosin may also be effective. Other sympatholytic agents, such as methyldopa, guanethidine, and phenoxybenzamine, may be useful in some patients. Surgical sympathectomy is helpful in some patients who are unresponsive to medical therapy, but benefit is often transient.

ACROCYANOSIS

In this condition, there is arterial vasoconstriction and secondary dilation of the capillaries and venules with resulting persistent cyanosis of the hands and, less frequently, the feet. Cyanosis may be intensified by exposure to a cold environment. Women are affected much more frequently than men, and the age of onset is usually less than 30 years. Generally, patients are asymptomatic but seek medical attention because of the discoloration. Examination reveals normal pulses, peripheral cyanosis, and moist palms. Trophic skin changes and ulcerations do *not* occur. The disorder can be distinguished from Raynaud's phenomenon because it is persistent and not episodic, the discoloration extends proximally from the digits, and blanching does not occur. Ischemia secondary to arterial occlusive disease can usually be excluded by the presence of normal pulses. Central cyanosis and decreased arterial oxygen saturation are not present. Patients should be reassured and advised to dress warmly and avoid cold exposure. Pharmacologic intervention is not indicated.

LIVEDO RETICULARIS

In this condition, localized areas of the extremities develop a mottled or netlike appearance of reddish to blue discoloration. The mottled appearance may be more prominent following cold exposure. The idiopathic form of this disorder occurs equally in men and women, and the most common age of onset is in the third decade. Patients with the idiopathic form are usually asymptomatic and seek attention for cosmetic reasons. Livedo reticularis can also occur following atheroembolism (see above). Rarely, skin ulcerations develop. Patients should be reassured and advised to avoid cold environments. No drug treatment is indicated.

PERNIO (CHILBLAINS)

This is a vasculitic disorder associated with exposure to cold; acute forms have been described. Raised erythematous lesions develop on the lower part of the legs and feet in cold weather. These are associated with pruritus and a burning sensation, and they may blister and ulcerate. Pathologic examination demonstrates angiitis characterized by intimal proliferation and perivascular infiltration of mononuclear and polymorphonuclear leukocytes. Giant cells may be present in the subcutaneous tissue. Patients should avoid exposure to cold, and ulcers should be kept clean and protected with sterile dressings. Sympatholytic drugs may be effective in some patients.

ERYTHROMELALGIA (ERYTHERMALGIA)

This disorder is characterized by burning pain and erythema of the extremities. The feet are involved more frequently than the hands, and males are affected more frequently

than females. Erythromelalgia may occur at any age but is most common in middle age. It may be primary or secondary to myeloproliferative disorders such as polycythemia vera and essential thrombocytosis, or it may occur as an adverse effect of drugs such as nifedipine or bromocriptine. Patients complain of burning in the extremities that is precipitated by exposure to a warm environment and aggravated by a dependent position. The symptoms are relieved by exposing the affected area to cool air or water or by elevation. Erythromelalgia can be distinguished from ischemia secondary to peripheral arterial disorders and peripheral neuropathy because the peripheral pulses are present and the neurologic examination is normal. There is no specific treatment; aspirin may produce relief in patients with erythromelalgia secondary to myeloproliferative disease. Treatment of associated disorders in secondary erythromelalgia may be helpful.

FROSTBITE

In this condition, tissue damage results from severe environmental cold exposure or from direct contact with a very cold object. Tissue injury results from both freezing and vasoconstriction. Frostbite usually affects the distal aspects of the extremities or exposed parts of the face, such as the ears, nose, chin, and cheeks. Superficial frostbite involves the skin and subcutaneous tissue. Patients experience pain or paresthesia, and the skin appears white and waxy. After rewarming, there is cyanosis and erythema, wheal- and-flare formation, edema, and superficial blisters. Deep frostbite involves muscle, nerves, and deeper blood vessels. It may result in edema of the hand or foot, vesicles and bullae, tissue necrosis, and gangrene.

Initial treatment is rewarming, performed in an environment where reexposure to freezing conditions will not occur. Rewarming is accomplished by immersion of the affected part in a water bath at temperatures of 40 to 44°C (104 to 111°F). Massage, application of ice water, and extreme heat are contraindicated. The injured area should be cleansed with soap or antiseptic and sterile dressings applied. Analgesics are often required during rewarming. Antibiotics are used if there is evidence of infection. The efficacy of sympathetic blocking drugs is not established. Following recovery, the affected extremity may exhibit increased sensitivity to cold.

VENOUS DISORDERS

Veins in the extremities can be broadly classified as either superficial or deep. In the lower extremity, the superficial venous system includes the greater and lesser saphenous veins and their tributaries. The deep veins of the leg accompany the major arteries. Perforating veins connect the superficial and deep systems at multiple locations. Bicuspid valves are present throughout the venous system to direct the flow of venous blood centrally.

VENOUS THROMBOSIS

The presence of thrombus within a superficial or deep vein and the accompanying inflammatory response in the vessel wall is termed *venous thrombosis* or *thrombophlebitis*. Initially, the thrombus is composed principally of platelets and fibrin. Red cells become interspersed with fibrin, and the thrombus tends to propagate in the

direction of blood flow. The inflammatory response in the vessel wall may be minimal or characterized by granulocyte infiltration, loss of endothelium, and edema.

The factors that predispose to venous thrombosis were initially described by Virchow in 1856 and include stasis, vascular damage, and hypercoagulability. Accordingly, a variety of clinical situations are associated with increased risk of venous thrombosis (Table 248-2). Venous thrombosis may occur in more than 50% of patients having orthopedic surgical procedures, particularly those involving the hip or knee, and in 10 to 40% of patients who undergo abdominal or thoracic operations. The prevalence of venous thrombosis is particularly high in patients with cancer of the pancreas, lungs, genitourinary tract, stomach, and breast. Approximately 10 to 20% of patients with idiopathic deep vein thrombosis have or develop clinically overt cancer; there is no consensus on whether these individuals should be subjected to intensive diagnostic workup to search for occult malignancy. Risk of thrombosis is increased following trauma, such as fractures of the spine, pelvis, femur, and tibia. Immobilization, regardless of the underlying disease, is a major predisposing cause of venous thrombosis. This fact may account for the relatively high incidence in patients with acute myocardial infarction or congestive heart failure. The incidence of venous thrombosis is increased during pregnancy, particularly in the third trimester and in the first month postpartum, and in individuals who use oral contraceptives or receive postmenopausal hormone replacement therapy. A variety of clinical disorders that produce systemic hypercoagulability, including resistance to activated protein C (factor V Leiden); antithrombin III, protein C, and protein S deficiencies; antiphospholipid syndrome;SLE; myeloproliferative diseases; dysfibrinogenemia; and disseminated intravascular coagulation, are associated with venous thrombosis. Venulitis occurring in thromboangiitis obliterans, Behcet's disease, and homocysteinuria may also cause venous thrombosis.

DEEP VENOUS THROMBOSIS

The most important consequences of this disorder are pulmonary embolism (Chap. 261) and the syndrome of chronic venous insufficiency. Deep venous thrombosis of the iliac, femoral, or popliteal veins is suggested by unilateral leg swelling, warmth, and erythema. Tenderness may be present along the course of the involved veins, and a cord may be palpable. There may be increased tissue turgor, distention of superficial veins, and the appearance of prominent venous collaterals. In some patients, deoxygenated hemoglobin in stagnant veins imparts a cyanotic hue to the limb, a condition called *phlegmasia cerulea dolens*. In markedly edematous legs, the interstitial tissue pressure may exceed the capillary perfusion pressure, causing pallor, a condition designated *phlegmasia alba dolens*.

The diagnosis of deep venous thrombosis of the calf is often difficult to make at the bedside. This is so because only one of multiple veins may be involved, allowing adequate venous return through the remaining patent vessels. The most common complaint is calf pain. Examination may reveal posterior calf tenderness, warmth, increased tissue turgor or modest swelling, and, rarely, a cord. Increased resistance or pain during dorsiflexion of the foot (Homans' sign) is an unreliable diagnostic sign.

Deep venous thrombosis occurs less frequently in the upper extremity than in the lower

extremity, but the incidence is increasing because of greater utilization of indwelling central venous catheters. The clinical features and complications are similar to those described for the leg.

Diagnosis The noninvasive test used most often to diagnose deep venous thrombosis is duplex venous ultrasonography (B-mode, i.e., two-dimensional, imaging, and pulse-wave Doppler interrogation). By imaging the deep veins, thrombus can be detected either by direct visualization or by inference when the vein does not collapse on compressive maneuvers. The Doppler ultrasound measures the velocity of blood flow in veins. This velocity is normally affected by respiration and by manual compression of the foot or calf. Flow abnormalities occur when deep venous obstruction is present. The positive predictive value of duplex venous ultrasonography approaches 95% for proximal deep vein thrombosis. In the calf, because calf veins are more difficult to visualize than proximal veins, the sensitivity of this technique is only 50 to 75%, although its specificity is 95%.

Impedance plethysmography measures changes in venous capacitance during physiologic maneuvers. Venous obstruction blunts the normal changes in venous capacitance that occur following inflation and deflation of a thigh cuff. The predictive value of this test for detecting occlusive thrombi in proximal veins is approximately 90%. However, it is much less sensitive for diagnosing deep venous thrombosis of the calves.

Magnetic resonance imaging (MRI) is another noninvasive means to detect deep vein thrombosis. Its diagnostic accuracy for assessing proximal deep vein thrombosis is similar to that of duplex ultrasonography. It is useful in patients with suspected thrombosis of the superior and inferior venae cavae or pelvic veins.

Deep venous thrombosis can also be diagnosed by venography. Contrast medium is injected into a superficial vein of the foot and directed to the deep system by the application of tourniquets. The presence of a filling defect or absence of filling of the deep veins is required to make the diagnosis.

Deep vein thrombosis must be differentiated from a variety of disorders that cause unilateral leg pain or swelling, including muscle rupture, trauma, or hemorrhage; a ruptured popliteal cyst; and lymphedema. It may be difficult to distinguish swelling caused by the postphlebitic syndrome from that due to acute recurrent deep venous thrombosis. Leg pain may also result from nerve compression, arthritis, tendinitis, fractures, and arterial occlusive disorders. A careful history and physical examination can usually determine the cause of these symptoms.

TREATMENT

Anticoagulants (See also Chap. 261) Prevention of pulmonary embolism is the most important reason for treating patients with deep vein thrombosis, since in the early stages the thrombus may be loose and poorly adherent to the vessel wall. Patients should be placed in bed, and the affected extremity should be elevated above the level of the heart until the edema and tenderness subside. Anticoagulants prevent thrombus propagation and allow the endogenous lytic system to operate. Initial therapy should include either unfractionated heparin or low-molecular-weight heparin. Unfractionated

heparin should be administered intravenously as an initial bolus of 7500 to 10,000 IU, followed by a continuous infusion of 1000 to 1500 IU/h. The rate of the heparin infusion should be adjusted so that the activated partial thromboplastin time (aPTT) is approximately twice the control value. Subcutaneous injection of heparin has been used as an alternative form of therapy. In fewer than 5% of patients, heparin therapy may cause thrombocytopenia. Infrequently, these patients develop arterial thrombosis and ischemia. Low-molecular-weight (4000 to 6000 Da) heparins are reported to be as effective as or better than conventional, unfractionated heparin in preventing extension or recurrence of venous thrombosis. Depending on the specific preparation, low-molecular-weight heparin is administered subcutaneously, in fixed doses, once or twice daily; for example, the dose of enoxaparin is 1 mg/kg subcutaneously bid. The incidence of thrombocytopenia is less with low-molecular-weight heparin than with conventional preparations. Hirudin, a direct thrombin inhibitor, may be used as initial anticoagulant therapy for patients in whom heparin is contraindicated because of heparin-induced thrombocytopenia. Warfarin is administered during the first week of treatment with heparin and may be started as early as the first day of heparin treatment if the aPTT is therapeutic. It is important to overlap heparin treatment with oral anticoagulant therapy for at least 4 to 5 days because the full anticoagulant effect of warfarin is delayed. The dose of warfarin should be adjusted to maintain the prothrombin time at an international normalized ratio (INR) of 2.0 to 3.0.

Anticoagulant treatment is indicated for patients with proximal deep vein thrombosis, since pulmonary embolism may occur in approximately 50% of untreated individuals. The use of anticoagulants for isolated deep vein thrombosis of the calf is controversial. However, approximately 20 to 30% of calf thrombi propagate to the thigh, thereby increasing the risk of pulmonary embolism. The overall incidence of pulmonary embolism in patients presenting initially with deep calf vein thrombosis is 5 to 20%. Also, isolated calf vein thrombosis has been identified as a cause of embolic stroke via a patent foramen ovale. Therefore, patients with calf vein thrombosis should either receive anticoagulants or be followed with serial noninvasive tests to determine whether proximal propagation has occurred. Anticoagulant treatment should be continued for at least 3 to 6 months for patients with acute idiopathic deep vein thrombosis and for those with a temporary risk factor for venous thrombosis to decrease the chance of recurrence. The duration of treatment is indefinite for patients with recurrent deep vein thrombosis and for those in whom associated causes, such as malignancy or hypercoagulability, have not been eliminated. If treatment with anticoagulants is contraindicated because of a bleeding diathesis or risk of hemorrhage, protection from pulmonary embolism can be achieved by mechanically interrupting the flow of blood through the inferior vena cava. Inferior vena cava plication generally has been replaced by percutaneous insertion of a filter.

Thrombolytics Thrombolytic drugs such as streptokinase, urokinase, and tPA may also be used, but there is no evidence that thrombolytic therapy is more effective than anticoagulants in preventing pulmonary embolism. However, early administration of thrombolytic drugs may accelerate clot lysis, preserve venous valves, and decrease the potential for developing postphlebitic syndrome.

Prophylaxis Prophylaxis should be considered in clinical situations where the risk of deep vein thrombosis is high. Low-dose unfractionated heparin (5000 units 2 h prior to

surgery and then 5000 units every 8 to 12 h postoperatively), warfarin, and external pneumatic compression are all useful. Low-dose heparin reduces the risk of deep vein thrombosis associated with thoracic and abdominal surgery and with prolonged bed rest. Low-molecular-weight heparins have been shown to prevent deep vein thrombosis in patients undergoing general or orthopedic surgery and in acutely ill medical patients. They are said to be more effective than conventional heparin and to cause an equal or lower incidence of bleeding. Danaparoid, a low-molecular-weight heparinoid, may be used for prophylaxis in patients undergoing hip surgery. Warfarin in a dose that yields a prothrombin time equivalent to anINR of 2.0 to 3.0 is effective in preventing deep vein thrombosis associated with bone fractures and orthopedic surgery. Warfarin is started the night before surgery and continued throughout the convalescent period. External pneumatic compression devices applied to the legs are used to prevent deep vein thrombosis when even low doses of heparin or warfarin might cause serious bleeding, as during neurosurgery or transurethral resection of the prostate.

SUPERFICIAL VEIN THROMBOSIS

Thrombosis of the greater or lesser saphenous veins or their tributaries -- i.e., superficial vein thrombosis -- does not result in pulmonary embolism. It is associated with intravenous catheters and infusions, occurs in varicose veins, and may develop in association with deep vein thrombosis. Migrating superficial vein thrombosis is often a marker for a carcinoma and may also occur in patients with vasculitides, such as thromboangiitis obliterans. The clinical features of superficial vein thrombosis are easily distinguished from those of deep vein thrombosis. Patients complain of pain localized to the site of the thrombus. Examination reveals a reddened, warm, and tender cord extending along a superficial vein. The surrounding area may be red and edematous.

TREATMENT

Treatment is primarily supportive. Initially, patients can be placed at bed rest with leg elevation and application of warm compresses. Nonsteroidal antiinflammatory drugs may provide analgesia but may also obscure clinical evidence of thrombus propagation. If a thrombosis of the greater saphenous vein develops in the thigh and extends toward the saphenofemoral vein junction, it is reasonable to consider anticoagulant therapy to prevent extension of the thrombus into the deep system and a possible pulmonary embolism.

VARICOSE VEINS

Varicose veins are dilated, tortuous superficial veins that result from defective structure and function of the valves of the saphenous veins, from intrinsic weakness of the vein wall, from high intraluminal pressure, or, rarely, from arteriovenous fistulas. Varicose veins can be categorized as primary or secondary. Primary varicose veins originate in the superficial system and occur two to three times as frequently in women as in men. Approximately half of patients have a family history of varicose veins. Secondary varicose veins result from deep venous insufficiency and incompetent perforating veins or from deep venous occlusion causing enlargement of superficial veins that are serving as collaterals.

Patients with venous varicosities are often concerned about the cosmetic appearance of their legs. Symptoms consist of a dull ache or pressure sensation in the legs after prolonged standing; it is relieved with leg elevation. The legs feel heavy, and mild ankle edema develops occasionally. Extensive venous varicosities may cause skin ulcerations near the ankle. Superficial venous thrombosis may be a recurring problem, and, rarely, a varicosity ruptures and bleeds. Visual inspection of the legs in the dependent position usually confirms the presence of varicose veins.

Varicose veins can usually be treated with conservative measures. Symptoms often decrease when the legs are elevated periodically, when prolonged standing is avoided, and when elastic support hose are worn. External compression stockings provide a counterbalance to the hydrostatic pressure in the veins. Small symptomatic varicose veins can be treated with sclerotherapy, in which a sclerosing solution is injected into the involved varicose vein and a compression bandage is applied. Surgical therapy usually involves extensive ligation and stripping of the greater and lesser saphenous veins and should be reserved for patients who are very symptomatic, suffer recurrent superficial vein thrombosis, and/or develop skin ulceration. Surgical therapy may also be indicated for cosmetic reasons.

CHRONIC VENOUS INSUFFICIENCY

Chronic venous insufficiency may result from deep vein thrombosis and/or valvular incompetence. Following deep vein thrombosis, the delicate valve leaflets become thickened and contracted so that they cannot prevent retrograde flow of blood; the vein becomes rigid and thick-walled. Although most veins recanalize after an episode of thrombosis, the large proximal veins may remain occluded. Secondary incompetence develops in distal valves because high pressures distend the vein and separate the leaflets. Primary deep venous valvular dysfunction may also occur without previous thrombosis. Patients with venous insufficiency often complain of a dull ache in the leg that worsens with prolonged standing and resolves with leg elevation. Examination demonstrates increased leg circumference, edema, and superficial varicose veins. Erythema, dermatitis, and hyperpigmentation develop along the distal aspect of the leg, and skin ulceration may occur near the medial and lateral malleoli. Cellulitis may be a recurring problem. Patients should be advised to avoid prolonged standing or sitting; frequent leg elevation is helpful. Graduated compression stockings should be worn during the day. These efforts should be intensified if skin ulcers develop. Ulcers should be treated with applications of wet to dry dressings and, occasionally, dilute topical antibiotic solutions. Commercially available dressings comprising antiseptic solutions and compressive bandages may be applied and should be changed weekly until healing occurs. Recurrent ulceration and severe edema may be treated by surgical interruption of incompetent communicating veins. Rarely, surgical valvuloplasty and bypass of venous occlusions are employed.

LYMPHATIC DISORDERS

Lymphatic capillaries are blind-ended tubes formed by a single layer of endothelial cells. The absent or widely fenestrated basement membrane of lymphatic capillaries allows access to interstitial proteins and particles. Lymphatic capillaries merge to form larger vessels which contain smooth muscle and are capable of vasomotion. Small and

medium-sized lymphatic vessels empty into progressively larger channels, most of which drain into the thoracic duct. The lymphatic circulation is involved in the absorption of interstitial fluid and in the response to infection.

LYMPHEDEMA

Lymphedema may be categorized as primary or secondary (Table 248-3). The prevalence of primary lymphedema is approximately 1 per 10,000 individuals. Primary lymphedema may be secondary to agenesis, hypoplasia, or obstruction of the lymphatic vessels. It may be associated with Turner syndrome, Klinefelter syndrome, Noonan syndrome, the yellow nail syndrome, the intestinal lymphangiectasia syndrome, and lymphangiomyomatosis. Women are affected more frequently than men. There are three clinical subtypes: congenital lymphedema, which appears shortly after birth; lymphedema praecox, which has its onset at the time of puberty; and lymphedema tarda, which usually begins after age 35. Familial forms of congenital lymphedema (Milroy's disease) and lymphedema praecox (Meige's disease) may be inherited in an autosomal dominant manner with variable penetrance; autosomal or sex-linked recessive forms are less common.

Secondary lymphedema is an acquired condition resulting from damage to or obstruction of previously normal lymphatic channels (Table 248-3). Recurrent episodes of bacterial lymphangitis, usually caused by streptococci, are a very common cause of lymphedema. The most common cause of secondary lymphedema worldwide is filariasis (Chap. 221). Tumors, such as prostate cancer and lymphoma, can also obstruct lymphatic vessels. Both surgery and radiation therapy for breast carcinoma may cause lymphedema of the upper extremity. Less common causes include tuberculosis, contact dermatitis, lymphogranuloma venereum, rheumatoid arthritis, pregnancy, and self-induced or factitious lymphedema following application of tourniquets.

Lymphedema is generally a painless condition, but patients may experience a chronic dull, heavy sensation in the leg, and most often they are concerned about the appearance of the leg. Lymphedema of the lower extremity, initially involving the foot, gradually progresses up the leg so that the entire limb becomes edematous. In the early stages, the edema is soft and pits easily with pressure. In the chronic stages, the limb has a woody texture, and the tissues become indurated and fibrotic. At this point the edema may no longer be pitting. The limb loses its normal contour, and the toes appear square. Lymphedema should be distinguished from other disorders that cause unilateral leg swelling, such as deep vein thrombosis and chronic venous insufficiency. In the latter condition, the edema is softer, and there is often evidence of a stasis dermatitis, hyperpigmentation, and superficial venous varicosities.

The evaluation of patients with lymphedema should include diagnostic studies to clarify the cause. Abdominal and pelvic ultrasound and computed tomography can be used to detect obstructing lesions such as neoplasms.MRI may reveal edema in the epifascial compartment and identify lymph nodes and enlarged lymphatic channels. Lymphoscintigraphy and lymphangiography are rarely indicated, but either can be used to confirm the diagnosis or to differentiate primary from secondary lymphedema. Lymphoscintigraphy involves the injection of radioactively labeled technetium-containing

colloid into the distal subcutaneous tissue of the affected extremity. In lymphangiography, contrast material is injected into a distal lymphatic vessel that has been isolated and cannulated. In primary lymphedema, lymphatic channels are absent, hypoplastic, or ectatic. In secondary lymphedema, lymphatic channels are usually dilated, and it may be possible to determine the level of obstruction.

TREATMENT

Patients with lymphedema of the lower extremities must be instructed to take meticulous care of their feet to prevent recurrent lymphangitis. Skin hygiene is important, and emollients can be used to prevent drying. Prophylactic antibiotics are often helpful, and fungal infection should be treated aggressively. Patients should be encouraged to participate in physical activity; frequent leg elevation can reduce the amount of edema. Physical therapy, including massage to facilitate lymphatic drainage, may be helpful. Patients can be fitted with graduated compression hose to reduce the amount of lymphedema that develops with upright posture. Occasionally, intermittent pneumatic compression devices can be applied at home to facilitate reduction of the edema. Diuretics are contraindicated and may cause depletion of intravascular volume and metabolic abnormalities. Recently, microsurgical lymphatico-venous anastomotic procedures have been performed to rechannel lymph flow from obstructed lymphatic vessels into the venous system.

(Bibliography omitted in Palm version)

Back to Table of Contents

PART NINE -DISORDERS OF THE RESPIRATORY SYSTEM

SECTION 1 -DIAGNOSIS

249. APPROACH TO THE PATIENT WITH DISEASE OF THE RESPIRATORY SYSTEM - *Jeffrey M. Drazen*, *Steven E. Weinberger*

Patients with disease of the respiratory system generally present because of symptoms, an abnormality on a chest radiograph, or both. A set of diagnostic possibilities is often suggested by the initial problems at presentation, including the particular symptom(s) and the appearance of any radiographic abnormalities. The differential diagnosis is then refined on the basis of additional information gleaned from physical examination, pulmonary function testing, additional imaging studies, and bronchoscopic examination. This**chapter**will consider the approach to the patient based on the major patterns of presentation, focusing on the history, the physical examination, and the chest radiograph.**For further discussion of pulmonary function testing, see Chap. 250, and of other diagnostic studies, see Chap. 251.*

CLINICAL PRESENTATION

HISTORY

Dyspnea (shortness of breath) and cough are the primary presenting symptoms for patients with respiratory system disease. Less common symptoms include hemoptysis (the coughing up of blood) and chest pain, often with a pleuritic quality.

Dyspnea (See alsoChap. 32) When evaluating a patient with shortness of breath, one should first determine the time course over which the symptom has become manifest. Patients who were well previously and developed *acute* shortness of breath (over a period of hours to days) can have acute disease affecting the airways (an acute attack of asthma), the pulmonary parenchyma (acute pulmonary edema or an acute infectious process such as a bacterial pneumonia), the pleural space (a pneumothorax), or the pulmonary vasculature (a pulmonary embolus). A *subacute* presentation (over days to weeks) can suggest an exacerbation of preexisting airways disease (asthma or chronic bronchitis), a parenchymal infection or a noninfectious inflammatory process that proceeds at a relatively slow pace (*Pneumocystis carinii* pneumonia in a patient with AIDS, mycobacterial or fungal pneumonia, Wegener's granulomatosis, eosinophilic pneumonia, bronchiolitis obliterans with organizing pneumonia, and many others), neuromuscular disease (Guillain-Barre syndrome, myasthenia gravis), pleural disease (pleural effusion from a variety of possible causes), or chronic cardiac disease (congestive heart failure). A *chronic* presentation (over months to years) often indicates chronic obstructive lung disease, chronic interstitial lung disease, or chronic cardiac disease. Chronic diseases of airways (not only chronic obstructive lung disease but also asthma) are characterized by exacerbations and remissions. Patients often have periods when they are severely limited by shortness of breath, but these may be interspersed with periods in which symptoms are minimal or absent. In contrast, many of the diseases of pulmonary parenchyma are characterized by a slow but inexorable progression.

Other Respiratory Symptoms *Cough* (Chap. 33) may indicate the presence of lung disease, but cough per se is not useful for the differential diagnosis. The presence of sputum accompanying the cough often suggests airway disease and may be seen in asthma, chronic bronchitis, or bronchiectasis.

Hemoptysis (Chap. 33) can originate from disease of the airways, the pulmonary parenchyma, or the vasculature. Diseases of the airways can be inflammatory (acute or chronic bronchitis, bronchiectasis, or cystic fibrosis) or neoplastic (bronchogenic carcinoma or bronchial carcinoid tumors). Parenchymal diseases causing hemoptysis may be either localized (pneumonia, lung abscess, tuberculosis, or infection with *Aspergillus*) or diffuse (Goodpasture's syndrome, idiopathic pulmonary hemosiderosis). Vascular diseases potentially associated with hemoptysis include pulmonary thromboembolic disease and pulmonary arteriovenous malformations.

Chest pain (Chap. 13) caused by diseases of the respiratory system usually originates from involvement of the parietal pleura. As a result, the pain is accentuated by respiratory motion and is often referred to as *pleuritic*. Common examples include primary pleural disorders, such as neoplasm or inflammatory disorders involving the pleura, or pulmonary parenchymal disorders that extend to the pleural surface, such as pneumonia or pulmonary infarction.

Additional Historic Information Information about risk factors for lung disease should be explicitly explored to assure a complete basis of historic data. A history of current and past smoking, especially of cigarettes, should be sought from all patients. The smoking history should include the number of years of smoking, the intensity (i.e., number of packs per day), and, if the patient no longer smokes, the interval since smoking cessation. The risk of lung cancer falls progressively with the interval following discontinuation of smoking, and loss of lung function above the expected age-related decline ceases with the discontinuation of smoking. Even though chronic obstructive lung disease and neoplasia are the two most important respiratory complications of smoking, other respiratory disorders (e.g., spontaneous pneumothorax, respiratory bronchiolitis-interstitial lung disease, eosinophilic granuloma of the lung, and pulmonary hemorrhage with Goodpasture's syndrome) are also associated with smoking. A history of significant secondhand (passive) exposure to smoke, whether in the home or at the workplace, should also be sought as it may be a risk factor for neoplasia or an exacerbating factor for airways disease.

The patient may have been exposed to other inhaled agents associated with lung disease, which act either via direct toxicity or through immune mechanisms (Chaps. 253 and254). Such exposures can be either occupational or avocational, indicating the importance of detailed occupational and personal histories, the latter stressing exposures related to hobbies or the home environment. Important agents include the inorganic dusts associated with pneumoconiosis (especially asbestos and silica dusts) and organic antigens associated with hypersensitivity pneumonitis (especially antigens from molds and animal proteins). Asthma, which is more common in women than men, is often exacerbated by exposure to environmental allergens (dust mites, pet dander, or cockroach allergens in the home or allergens in the outdoor environment such as pollen and ragweed) or may be caused by occupational exposures (diisocyanates). Exposure to particular infectious agents can be suggested by contacts with individuals with known

respiratory infections (especially tuberculosis) or by residence in an area with endemic pathogens (histoplasmosis, coccidioidomycosis, blastomycosis).

A history of coexisting nonrespiratory disease or of risk factors for or previous treatment of such diseases should be sought, as they may predispose a patient to both infectious and noninfectious respiratory system complications. Common examples include systemic rheumatic diseases that are associated with pleural or parenchymal lung disease (Chap. 312), metastatic neoplastic disease in the lung, or impaired host defense mechanisms and secondary infection, which occur in the case of hematologic and lymph node malignancies. Risk factors for AIDS should be sought, as the lungs are not only the most common site of AIDS-defining infection but also can be involved by nonfectious complications of AIDS (Chap. 309). Treatment of nonrespiratory disease can be associated with respiratory complications, either because of effects on host defense mechanisms (immunosuppressive agents, cancer chemotherapy) with resulting infection or because of direct effects on the pulmonary parenchyma (cancer chemotherapy, radiation therapy, or treatment with other agents, such as amiodarone) or on the airways (beta-blocking agents causing airflow obstruction, angiotensin-converting enzyme inhibitors causing cough) (Chap. 253).

Family history is important for evaluating diseases that have a genetic component. These include disorders such as cystic fibrosis,a_1-antitrypsin deficiency, and asthma.

PHYSICAL EXAMINATION

The general principles of inspection, palpation, percussion, and auscultation apply to the examination of the respiratory system. However, the physical examination should be directed not only toward ascertaining abnormalities of the lungs and thorax but also toward recognizing other findings that may reflect underlying lung disease.

On *inspection*, the rate and pattern of breathing as well as the depth and symmetry of lung expansion are observed. Breathing that is unusually rapid, labored, or associated with the use of accessory muscles of respiration generally indicates either augmented respiratory demands or an increased work of breathing. Asymmetric expansion of the chest is usually due to an asymmetric process affecting the lungs, such as endobronchial obstruction of a large airway, unilateral parenchymal or pleural disease, or unilateral phrenic nerve paralysis. Visible abnormalities of the thoracic cage include kyphoscoliosis and ankylosing spondylitis, either of which can alter compliance of the thorax, increase the work of breathing, and cause dyspnea.

On *palpation*, the symmetry of lung expansion can be assessed, generally confirming the findings observed by inspection. Vibration produced by spoken sounds is transmitted to the chest wall and is assessed by the presence or absence and symmetry of tactile fremitus. Transmission of vibration is decreased or absent if pleural liquid is interposed between the lung and the chest wall or if an endobronchial obstruction alters sound transmission. In contrast, transmitted vibration may increase over an area of underlying pulmonary consolidation.

The relative resonance or dullness of the tissue underlying the chest wall is assessed by *percussion*. The normal sound of underlying air-containing lung is resonant. In contrast,

consolidated lung or a pleural effusion sounds dull, while emphysema or air in the pleural space results in a hyperresonant percussion note.

On *auscultation* of the lungs, the examiner listens for both the quality and intensity of the breath sounds and for the presence of extra, or adventitious, sounds. Normal breath sounds heard through the stethoscope at the periphery of the lung are described as *vesicular breath sounds*, in which inspiration is louder and longer than expiration. If sound transmission is impaired by endobronchial obstruction or by air or liquid in the pleural space, breath sounds are diminished in intensity or absent. When sound transmission is improved through consolidated lung, the resulting *bronchial breath sounds* have a more tubular quality and a more pronounced expiratory phase. Sound transmission can also be assessed by listening to spoken or whispered sounds; when these are transmitted through consolidated lung, *bronchophony* and *whispered pectoriloquy*, respectively, are present. The sound of a spoken E becomes more like an A, though with a nasal or bleating quality, a finding that is termed *egophony*.

The primary adventitious (abnormal) sounds that can be heard include crackles (rales), wheezes, and rhonchi. *Crackles* represent the typically inspiratory sound created when alveoli and small airways open and close with respiration, and they are often associated with interstitial lung disease, microatelectasis, or filling of alveoli by liquid. *Wheezes*, which are generally more prominent during expiration than inspiration, reflect the oscillation of airway walls that occurs when there is airflow limitation, as may be produced by bronchospasm, airway edema or collapse, or intraluminal obstruction by neoplasm or secretions. *Rhonchi* is the term applied to the sounds created when there is free liquid in the airway lumen; the viscous interaction between the free liquid and the moving air creates a low-pitched vibratory sound. Other adventitious sounds include pleural friction rubs and stridor. The gritty sound of a *pleural friction rub* indicates inflamed pleural surfaces rubbing against each other, often during both inspiratory and expiratory phases of the respiratory cycle. *Stridor*, which occurs primarily during inspiration, represents flow through a narrowed upper airway, as occurs in an infant with croup.

A summary of the patterns of physical findings on pulmonary examination in common types of respiratory system disease is shown in Table 249-1.

A meticulous *general physical examination* is mandatory in patients with disorders of the respiratory system. Enlarged lymph nodes in the cervical and supraclavicular regions should be sought. Disturbances of mentation or even coma can occur in patients with acute carbon dioxide retention and hypoxemia. Telltale stains on the fingers point to heavy cigarette smoking; infected teeth and gums may occur in patients with aspiration pneumonitis and lung abscess.

Clubbing of the digits can be found in lung cancer, interstitial lung disease, and chronic infections in the thorax, such as bronchiectasis, lung abscess, and empyema. Clubbing can also be seen with congenital heart disease associated with right-to-left shunting and with a variety of chronic inflammatory or infectious diseases, such as inflammatory bowel disease and endocarditis. A number of systemic diseases, such as systemic lupus erythematosus, scleroderma, and rheumatoid arthritis, may be associated with pulmonary complications, even though their primary clinical manifestations and physical

findings are not primarily related to the lungs. Conversely, other diseases that most commonly affect the respiratory system, such as sarcoidosis, can have findings on physical examination not related to the respiratory system, including ocular findings (uveitis, conjunctival granulomas) and skin findings (erythema nodosum, cutaneous granulomas).

CHEST RADIOGRAPHY

Chest radiography is often the initial diagnostic study performed to evaluate patients with respiratory symptoms, but it can also provide the initial evidence of disease in patients who are free of symptoms. Perhaps the most common example of the latter situation is the finding of one or more nodules or masses when the radiograph is performed for a reason other than evaluation of respiratory symptoms.

A number of diagnostic possibilities are often suggested by the radiographic pattern (Figs. 249-1 and249-2). A localized region of opacification involving the pulmonary parenchyma can be described as a nodule (usually <6 cm in diameter), a mass (usually ^{3} 6 cm in diameter), or an infiltrate. Diffuse disease with increased opacification is usually characterized as having an alveolar, an interstitial, or a nodular pattern. In contrast, increased radiolucency can be localized, as seen with a cyst or bulla, or generalized, as occurs with emphysema. The chest radiograph is also particularly useful for the detection of pleural disease, especially if manifested by the presence of air or liquid in the pleural space. An abnormal appearance of the hila and/or the mediastinum can suggest a mass or enlargement of lymph nodes.

A summary of representative diagnoses suggested by these common radiographic patterns is presented inTable 249-2.

Additional Diagnostic Evaluation Further information for clarification of radiographic abnormalities is frequently obtained with computed tomographic scanning of the chest (Chap. 251; seeFig. 265-2). This technique is more sensitive than plain radiography in detecting subtle abnormalities and can suggest specific diagnoses based on the pattern of abnormality.**For further discussion of the use of other imaging studies, including magnetic resonance imaging, scintigraphic studies, ultrasound, and angiography, see Chap. 251.*

Alteration in the function of the lungs as a result of respiratory system disease is assessed objectively by pulmonary function tests, and effects on gas exchange are evaluated by measurement of arterial blood gases or by oximetry (Chap. 250). As part of pulmonary function testing, quantitation of forced expiratory flow assesses the presence of obstructive physiology, which is consistent with diseases affecting the structure or function of the airways, such as asthma and chronic obstructive lung disease. Measurement of lung volumes assesses the presence of restrictive disorders, seen with diseases of the pulmonary parenchyma or respiratory pump and with space-occupying processes within the pleura.

Bronchoscopy is useful in some settings for visualizing abnormalities of the airways and for obtaining a variety of samples from either the airway or the pulmonary parenchyma (Chap. 251).

INTEGRATION OF THE PRESENTING CLINICAL PATTERN AND DIAGNOSTIC STUDIES

Patients with respiratory symptoms but a normal chest radiograph most commonly have diseases affecting the airways, such as asthma or chronic obstructive pulmonary disease. However, the latter diagnosis is also commonly associated with radiographic abnormalities, such as diaphragmatic flattening and attenuation of vascular markings. Other disorders of the respiratory system for which the chest radiograph is normal include disorders of the respiratory pump (either the chest wall or the neuromuscular apparatus controlling the chest wall) or pulmonary circulation and occasionally interstitial lung disease. Chest examination and pulmonary function tests are generally helpful in sorting out these diagnostic possibilities. Obstructive diseases associated with a normal or relatively normal chest radiograph are often characterized by findings on physical examination and pulmonary function testing that are typical for these conditions. Similarly, diseases of the respiratory pump or interstitial diseases may also be suggested by findings on physical examination or by particular patterns of restrictive disease seen on pulmonary function testing.

When respiratory symptoms are accompanied by radiographic abnormalities, diseases of the pulmonary parenchyma or the pleura are usually present. Either diffuse or localized parenchymal lung disease is generally visualized well on the radiograph, and both air and liquid in the pleural space (pneumothorax and pleural effusion, respectively) are usually readily detected by radiography.

Radiographic findings in the absence of respiratory symptoms often indicate localized disease affecting the airways or the pulmonary parenchyma. One or more nodules or masses can suggest intrathoracic malignancy, but they also can be the manifestation of a current or previous infectious process. Patients with diffuse parenchymal lung disease on radiographic examination may be free of symptoms, as is sometimes the case with pulmonary sarcoidosis.

In approaching the patient with pulmonary disease, consideration must be given to the observation that substantial changes in the relative incidence of diseases affecting the respiratory system have taken place in the United States during the past four decades. The prevalence of chronic infectious disorders such as lung abscess and bronchiectasis has decreased. Tuberculosis declined only to undergo resurgence when two susceptible populations, patients with AIDS and immigrants from Southeast Asia, increased in number. Patients with chronic bronchitis and with emphysema now survive longer and form an increasing fraction of patients with chronic respiratory disease, as do patients with environmental lung disease and with drug-induced pulmonary disease. Modern intercontinental travel has increased the appearance in the western world of parasitic infestations of the lung. Also, the reduction of immune competence that occurs in patients with AIDS and in those with diabetes as well as in patients being treated for a variety of malignancies and those receiving immunosuppressive drugs has led to an increasing incidence of opportunistic infections of the lungs with a variety of microorganisms that were rarely pathogenic in the past.

(Bibliography omitted in Palm version)

Back to Table of Contents

250. DISTURBANCES OF RESPIRATORY FUNCTION - *Steven E. Weinberger, Jeffrey M. Drazen*

The respiratory system includes the lungs, the central nervous system (CNS), the chest wall (with the diaphragm and intercostal muscles), and the pulmonary circulation. The CNS controls the activity of the muscles of the chest wall, which constitute the pump of the respiratory system. Because these components of the respiratory system act in concert to achieve gas exchange, malfunction of an individual component or alteration of the relationships among components can lead to disturbances in function. In this **chapter** we consider three major aspects of disturbed respiratory function: (1) disturbances in ventilatory function, (2) disturbances in the pulmonary circulation, and (3) disturbances in gas exchange. **For further discussion of disorders relating to CNS control of ventilation, see Chap. 263.*

DISTURBANCES IN VENTILATORY FUNCTION

Ventilation is the process whereby the lungs replenish the gas in the alveoli. Measurements of ventilatory function in common diagnostic use consist of quantification of the gas volume contained in the lungs under certain circumstances and the rate at which gas can be expelled from the lungs. Two measurements of lung volume commonly used for respiratory diagnosis are total lung capacity (TLC) and residual volume (RV). The former is the volume of gas contained in the lungs after a maximal inspiration, whereas the latter is the volume of gas remaining in the lungs at the end of a maximal expiration. The volume of gas that is exhaled from the lungs in going from TLC to RV is called the *vital capacity* (VC) (Fig. 250-1).

Common clinical measurements of airflow are obtained from maneuvers in which the subject inspires to TLC and then forcibly exhales to RV. Three measurements are commonly made from a recording of exhaled volume versus time -- i.e., a spirogram -- obtained during such a forced expiratory maneuver: (1) the volume of gas exhaled during the first second of expiration [forced expiratory volume (FEV) in 1 s, or FEV_1], (2) the total volume exhaled [forced vital capacity (FVC)], and (3) the average expiratory flow rate during the middle 50% of the VC [forced expiratory flow (FEF) between 25 and 75% of the VC, or $FEF_{25-75\%}$, also called the maximal midexpiratory flow rate (MMFR)] (Fig. 250-2).

PHYSIOLOGIC FEATURES

The lungs are elastic structures, containing collagen and elastic fibers that resist expansion. For normal lungs to contain air, they must be distended either by a positive internal pressure -- i.e., by a pressure in the airways and alveolar spaces -- or by a negative external pressure -- i.e., by a pressure outside the lung. The relationship between the volume of gas contained in the lungs and the distending pressure (the *transpulmonary pressure*, or P_{TP}, defined as internal pressure minus external pressure) is described by the pressure-volume curve of the lungs (Fig. 250-3*A*).

The chest wall is also an elastic structure, with properties similar to those of an expandable and compressible spring. The relationship between the volume enclosed by the chest wall and the distending pressure for the chest wall is described by the

pressure-volume curve of the chest wall (Fig. 250-3*B*). For the chest wall to assume a volume different from its resting volume, the internal or external pressures acting on it must be altered.

At functional residual capacity (FRC), defined as the volume of gas in the lungs at the end of a normal exhalation, the lungs are partially inflated, so their elastic recoil exerts a force tending to empty the lungs. At the same time, chest wall volume is such that its elastic recoil promotes outward expansion. FRC occurs at the lung volume at which the tendency of the lungs to contract is opposed by the equal and opposite tendency of the chest wall to expand (Fig. 250-3*C*).

For the lungs and the chest wall to achieve a volume other than the resting volume (FRC), either the pressures acting on them must be changed passively -- e.g., by a mechanical ventilator that delivers positive pressure to the airways and alveoli -- or the respiratory muscles must actively oppose the tendency of the lungs and the chest wall to return to FRC. During inhalation to volumes above FRC, the inspiratory muscles actively overcome the tendency of the respiratory system to decrease volume back to FRC. During active exhalation to volumes below FRC, expiratory muscle activity must overcome the tendency of the respiratory system to increase volume back to FRC.

AtTLC, the maximal force applied by the inspiratory muscles to expand the lungs is opposed mainly by the inward recoil of the lungs. As a consequence, the major determinants of TLC are the stiffness of the lungs and inspiratory muscle strength. If the lungs become stiffer -- i.e., less compliant -- TLC is decreased. If the lungs become less stiff (more compliant), TLC is increased. If the inspiratory muscles are significantly weakened, they are less able to overcome the inward elastic recoil of the lungs, and TLC is lowered.

AtRV, the force exerted by the expiratory muscles to decrease lung volume further is balanced by the outward recoil of the chest wall, which becomes extremely stiff at low lung volumes. Two factors influence the volume of gas contained in the lungs at RV. The first is the ability of the subject to exert a prolonged expiratory effort, which is related to muscle strength and the ability to overcome sensory stimuli from the chest wall. The second is the ability of the lungs to empty to a small volume. In normal lungs, as P_{TP} is lowered, lung volume decreases. In lungs with diseased airways, as P_{TP} is lowered, flow limitation or airway closure may limit the amount of gas that can be expired. Consequently, either weak expiratory muscles or intrinsic airways disease can result in an elevation in measured RV.

Dynamic measurements of ventilatory function are made by having the subject inhale toTLC and then perform a forced expiration toRV. If a subject performs a series of such expiratory maneuvers using increasing muscular intensity, expiratory flow rates will increase until a certain level of effort is reached. Beyond this level, additional effort at any given lung volume will not increase the forced expiratory flow rate; this phenomenon is known as the *effort independence* of forced expiratory flow. The physiologic mechanisms determining the flow rates during this effort-independent phase of FEF have been shown to be the elastic recoil of the lung, the airflow resistance of the airways between the alveolar zone and the physical site of flow limitation, and the airway wall compliance at the site of flow limitation. Physical processes that decrease

elastic recoil, increase airflow resistance, or increase airway wall compliance decrease the flow rate that can be achieved at any given lung volume. Conversely, processes that increase elastic recoil, decrease resistance, or stiffen airway walls increase the flow rate that can be achieved at any given lung volume.

MEASUREMENT OF VENTILATORY FUNCTION

Ventilatory function is measured under static conditions for determination of lung volumes and under dynamic conditions for determination of forced expiratory flow rates.VC, expiratory reserve volume (ERV), and inspiratory capacity (IC) (Fig. 250-1) are measured by having the patient breathe into and out of a spirometer, a device capable of measuring expired or inspired gas volume while plotting volume as a function of time. Other volumes -- specifically,RV,FRC, andTLC -- cannot be measured in this way because they include the volume of gas present in the lungs even after a maximal expiration. Two techniques are commonly used to measure these volumes: helium dilution and body plethysmography. In the helium dilution method, the subject repeatedly breathes in and out from a reservoir with a known volume of gas containing a trace amount of helium. The helium is diluted by the gas previously present in the lungs and very little is absorbed into the pulmonary circulation. From knowledge of the reservoir volume and the initial and final helium concentrations, the volume of gas present in the lungs can be calculated. The helium dilution method may underestimate the volume of gas in the lungs if there are slowly communicating airspaces, such as bullae. In this situation, lung volumes can be measured more accurately with a body plethysmograph, a sealed box in which the patient sits while panting against a closed mouthpiece. Because there is no airflow into or out of the plethysmograph, the pressure changes in the thorax during panting cause compression and rarefaction of gas in the lungs and simultaneous rarefaction and compression of gas in the plethysmograph. By measuring the pressure changes in the plethysmograph and at the mouthpiece, the volume of gas in the thorax can be calculated using Boyle's law.

Lung volumes and measurements made during forced expiration are interpreted by comparing the values measured with the values expected given the age, height, sex, and race of the patient (Appendix A). Regression curves have been constructed on the basis of data obtained from large numbers of normal, nonsmoking individuals without evidence of lung disease. Predicted values for a given patient can then be obtained by using the patient's age and height in the appropriate regression equation; different equations are used depending on the patient's race and gender. Because there is some variability among normal individuals, values between 80 and 120% of the predicted value have traditionally been considered normal. Increasingly, calculated percentiles are used in determining normality. Specifically, values of individual measurements falling below the fifth percentile are considered to be below normal.

The normal value for the ratioFEV_1/FVCis approximately 0.75 to 0.80, although this value does fall somewhat with advancing age. The$FEF_{25-75\%}$is often considered a more sensitive measurement of early airflow obstruction, particularly in small airways. However, this measurement must be interpreted cautiously in patients with abnormally small lungs (lowTLC andVC). These patients exhale less air during forced expiration, and the $FEF_{25-75\%}$may appear abnormal relative to the usual predicted value, even though it is normal relative to the size of the patient's lungs.

It is also a common practice to plot expiratory flow rates against lung volume (rather than against time); the close linkage of flow rates to lung volumes produces a typical *flow-volume curve* (Fig. 250-4). In addition, the spirometric values mentioned above can be calculated from the flow-volume curve. Commonly, flow rates during a maximal inspiratory effort performed as rapidly as possible are plotted as well, making the flow-volume curve into a *flow-volume loop*. AtTLC, before expiratory flow starts, the flow rate is zero; once forced expiration has begun, a high peak flow rate is rapidly achieved. As expiration continues and lung volume approachesRV, the flow rate falls progressively, in a nearly linear fashion as a function of lung volume for a person with normal lung function. During maximal inspiration from RV to TLC, inspiratory flow is most rapid at the midpoint of inspiration, so the inspiratory portion of the loop is U-shaped or saddle-shaped. The flow rates achieved during maximal expiration can be analyzed quantitatively by comparing the flow rates at specified lung volumes with the predicted values or qualitatively by analyzing the shape of the descending limb of the expiratory curve.

Assessing the strength of respiratory muscles is an additional part of the overall evaluation of some patients with respiratory dysfunction. When a patient exhales completely toRV and then tries to inspire maximally against an occluded airway, the pressure that can be generated is called the *maximal inspiratory pressure* (MIP). On the other hand, when a patient inhales toTLC and then tries to expire maximally against an occluded airway, the pressure generated is called the *maximal expiratory pressure* (MEP). In the proper clinical setting, these studies may provide useful information regarding the cause of abnormal lung volumes and the possibility that respiratory muscle weakness may be causally related to the lung volume abnormalities.

PATTERNS OF ABNORMAL FUNCTION

The two major patterns of abnormal ventilatory function, as measured by static lung volumes and spirometry, are restrictive and obstructive patterns. In the *obstructive pattern*, the hallmark is a decrease in expiratory flow rates. With fully established disease, the ratioFEV_1/FVCis decreased, as is the$FEF_{25-75\%}$(Fig. 250-2, line *B*). The expiratory portion of the flow-volume loop demonstrates decreased flow rates for any given lung volume. Nonuniform emptying of airways is reflected by a coved (concave upward) configuration of the curve (Fig. 250-4). With early obstructive disease, which originates in the small airways, FEV_1/FVC may be normal; the only abnormalities noted on routine testing of pulmonary function may be a depression in $FEF_{25-75\%}$and an abnormal, i.e., coved, configuration in the terminal portion of the forced expiratory flow-volume curve.

In *obstructive* disease, theTLC is normal or increased. When helium equilibration tests are used to measure lung volumes, the measured volume may be less than the actual volume if helium was not well distributed to all regions of the lung. Residual volume is elevated as a result of airway closure during expiration, and the ratioRV/TLC is increased.VC is frequently decreased in obstructive disease because of the striking elevations in RV with only minor changes in TLC.

A *restrictive pattern* can be broadly divided into two subgroups, depending on the

location of the pathology: pulmonary parenchymal and extraparenchymal. For extraparenchymal disease, dysfunction can be predominantly in inspiration or in both inspiration and expiration (Table 250-1). The hallmark of a restrictive pattern, found in all these subcategories, is a decrease in lung volumes, primarilyTLC andVC. In pulmonary parenchymal disease,RV is also generally decreased, and forced expiratory flow rates are preserved. In fact, when$FEV_1$is considered as a percentage of theFVC, the flow rates are often supranormal, i.e., disproportionately high relative to the size of the lungs (Fig. 250-2, line *C*). The flow-volume curve may graphically demonstrate this disproportionate relationship between flow rates and lung volumes, since the expiratory portion of the curve appears relatively tall (preserved flow rates) but narrow (decreased lung volumes), as shown inFig. 250-4.

In the extraparenchymal pattern characterized by *inspiratory dysfunction*, caused by either inspiratory muscle weakness or a stiff chest wall, inadequate distending forces are exerted on an otherwise normal lung. As a result,TLCvalues are less than predicted,RV is often not significantly affected, and expiratory flow rates are preserved. If inspiratory muscle weakness is the cause of this pattern, thenMIP is decreased. In the extraparenchymal pattern characterized by *inspiratory and expiratory dysfunction*, the ability to expire to a normal RV is also limited, because of either expiratory muscle weakness or a deformed chest wall that is abnormally rigid at volumes belowFRC. Consequently, RV is often elevated, unlike the pattern observed in the other restrictive subcategories. The ratioFEV_1/FVCis variable and depends on expiratory muscle strength. If expiratory muscle strength is significantly decreased, thenMEP is decreased, the ability to expire rapidly is impaired, and FEV_1/FVC may be decreased even though there is no airflow obstruction. If expiratory muscle strength is normal but the chest wall is abnormally stiff below FRC, then FEV_1/FVC is normal or increased.

CLINICAL CORRELATIONS

Table 250-1 summarizes the expected alterations in ventilatory function as indicated by pulmonary function testing. One reason to establish a ventilatory diagnosis is to categorize the functional disorder. This information can be useful in diagnosis, as outlined inTable 250-2. Note that lung disease such as pulmonary vascular disease or lung nodules can be present without abnormal ventilatory function, but the presence of specific diagnostic findings is an aid in differential diagnosis.

DISTURBANCES IN THE PULMONARY CIRCULATION

PHYSIOLOGIC FEATURES

The pulmonary vasculature must handle the entire output of the right ventricle, approximately 5 L/min in a normal adult at rest. The comparatively thin-walled vessels of the pulmonary arterial system provide relatively little resistance to flow and are capable of handling this large volume of blood at perfusion pressures that are low compared with those of the systemic circulation. The normal mean pulmonary artery pressure is 15 mmHg, as compared to approximately 95 mmHg for the normal mean aortic pressure. Regional blood flow in the lung is dependent on hydrostatic forces. In an upright person, pulmonary arterial pressure (PAP) is lowest at the apex of the lung and highest at the lung base. As a result, in the upright position, perfusion is least at the apex and greatest

at the base. When cardiac output increases, as occurs during exercise, the pulmonary vasculature is capable of recruiting previously unperfused vessels and distending underperfused vessels, thus responding to the increase in flow with a decrease in pulmonary vascular resistance. In consequence, the increase in mean PAP, even with a three- to fourfold increase in cardiac output, is small.

METHODS OF MEASUREMENT

Assessment of circulatory function in the pulmonary vasculature depends on measuring pulmonary vascular pressures and cardiac output. Clinically, these measurements are commonly made in intensive care units capable of invasive monitoring and in cardiac catheterization laboratories. With a flow-directed pulmonary arterial (Swan-Ganz) catheter,PAP and pulmonary capillary wedge pressure can be measured directly, and cardiac output can be obtained by the thermodilution method. Pulmonary vascular resistance (PVR) can then be calculated according to the equation

where PVR= pulmonary vascular resistance (dyn×s/cm^5); PAP = mean pulmonary arterial pressure (mmHg); PCW= pulmonary capillary wedge pressure (mmHg); and CO= cardiac output (L/min).

The normal value for pulmonary vascular resistance is approximately 50 to 150 dyn×s/cm^5.

MECHANISMS OF ABNORMAL FUNCTION (See alsoChap. 260)

PVRmay increase by a variety of mechanisms. Pulmonary arterial and arteriolar vasoconstriction is a prominent response to alveolar hypoxia. PVR also increases if intraluminal thrombi or proliferation of smooth muscle in vessel walls diminishes the luminal cross-sectional area. If small pulmonary vessels are destroyed, either by scarring or by loss of alveolar walls, the total cross-sectional area of the pulmonary vascular bed diminishes, and PVR increases. When PVR is elevated, eitherPAPrises to maintain normal cardiac output or cardiac output falls if PAP does not increase.

CLINICAL CORRELATIONS

Disturbances in the function of the pulmonary vasculature as a result of primary cardiac disease, either congenital heart disease or conditions that elevate left atrial pressure, such as mitral stenosis, are beyond the scope of this**chapter**and are discussed in Chaps. 234 and236, respectively. Instead, the focus will be on the pulmonary vasculature as its function is affected by diseases primarily involving the respiratory system, including the pulmonary vessels themselves.

All diseases of the respiratory system causing hypoxemia are potentially capable of increasingPVR, since alveolar hypoxia is a very potent stimulus for pulmonary vasoconstriction. The more prolonged and intense the hypoxic stimulus, the more likely it is that a significant increase in PVR producing pulmonary hypertension will result. In practice, patients with hypoxemia caused by chronic obstructive lung disease, interstitial

lung disease, chest wall disease, and the obesity hypoventilation-sleep apnea syndrome are particularly prone to developing pulmonary hypertension. If there are additional structural changes in the pulmonary vasculature secondary to the underlying process, these will increase the likelihood of developing pulmonary hypertension.

With diseases directly affecting the pulmonary vessels, a decrease in the cross-sectional area of the pulmonary vascular bed is primarily responsible for increasedPVR, while hypoxemia generally plays a lesser role. In the case of recurrent pulmonary emboli, parts of the pulmonary arterial system are occluded by intraluminal thrombi originating in the systemic venous system. With primary pulmonary hypertension (Chap. 260) or with pulmonary vascular disease secondary to scleroderma, the small pulmonary arteries and arterioles are affected by a generalized obliterative process that narrows and occludes these vessels. PVR increases, and significant pulmonary hypertension often results.

DISTURBANCES IN GAS EXCHANGE

PHYSIOLOGIC FEATURES

The primary functions of the respiratory system are to remove the appropriate amount of CO_2 from blood entering the pulmonary circulation and to provide adequate O_2 to blood leaving the pulmonary circulation. For these functions to be carried out properly, there must be adequate provision of fresh air to the alveoli for delivery of O_2 and removal of CO_2(ventilation), adequate circulation of blood through the pulmonary vasculature (perfusion), adequate movement of gas between alveoli and pulmonary capillaries (diffusion), and appropriate contact between alveolar gas and pulmonary capillary blood (ventilation-perfusion matching).

A normal individual at rest inspires approximately 12 to 16 times per minute, each breath having a tidal volume of approximately 500 mL. A portion (approximately 30%) of the fresh air inspired with each breath does not reach the alveoli but remains in the conducting airways of the lung. This component of each breath, which is not generally available for gas exchange, is called the *anatomic dead space component*. The remaining 70% reaches the alveolar zone, mixes rapidly with the gas already there, and can participate in gas exchange. In this example, the total ventilation each minute is approximately 7 L, composed of 2 L/min of dead space ventilation and 5 L/min of alveolar ventilation. In certain diseases, some alveoli are ventilated but not perfused, so that some ventilation in addition to the anatomic dead space component is wasted. If total dead space ventilation is increased but total minute ventilation is unchanged, then alveolar ventilation must fall correspondingly.

Gas exchange is dependent on alveolar ventilation rather than total minute ventilation, as outlined below. The partial pressure of CO_2 in arterial blood (Pa_{CO_2}) is directly proportional to the amount of CO_2produced per minute ($_{CO_2}$) and inversely proportional to alveolar ventilation (A), according to the relationship

where$_{CO_2}$ is expressed in mL/min, A in L/min, and Pa_{CO_2} in mmHg. At fixed$_{CO_2}$, when

alveolar ventilation increases, Pa_{CO_2} falls, and when alveolar ventilation decreases, Pa_{CO_2} rises. Maintaining a normal level of O_2 in the alveoli (and consequently in arterial blood) also depends on provision of adequate alveolar ventilation to replenish alveolar O_2. This principle will become more apparent from consideration of the alveolar gas equation below.

Diffusion of O_2 and CO_2 Both O_2 and CO_2 diffuse readily down their respective concentration gradients through the alveolar wall and pulmonary capillary endothelium. Under normal circumstances, this process is rapid, and equilibration of both gases is complete within one-third of the transit time of erythrocytes through the pulmonary capillary bed. Even in disease states in which diffusion of gases is impaired, the impairment is unlikely to be severe enough to prevent equilibration of CO_2 and O_2. Consequently, a diffusion abnormality rarely results in arterial hypoxemia at rest. If erythrocyte transit time in the pulmonary circulation is shortened, as occurs with exercise, and diffusion is impaired, then diffusion limitation may contribute to hypoxemia. Exercise testing can often demonstrate such physiologically significant abnormalities due to impaired diffusion. Even though diffusion limitation rarely makes a clinically significant contribution to resting hypoxemia, clinical measurements of what is known as *diffusing capacity* (see below) can be a useful measure of the integrity of the alveolar-capillary membrane.

Ventilation-Perfusion Matching In addition to the absolute levels of alveolar ventilation and perfusion, gas exchange depends critically on the proper matching of ventilation and perfusion. The spectrum of possible ventilation-perfusion (/) ratios in an alveolar-capillary unit ranges from zero, in which ventilation is totally absent and the unit behaves as a shunt, to infinity, in which perfusion is totally absent and the unit behaves as dead space. The P_{O_2} and P_{CO_2} of blood leaving each alveolar-capillary unit depend on the gas tension (of blood and air) entering that unit and on the particular / ratio of the unit. At one extreme, when an alveolar-capillary unit has a / ratio of 0 and behaves as a shunt, blood leaving the unit has the composition of mixed venous blood entering the pulmonary capillaries, i.e., P_{O_2}» 40 mmHg and P_{CO_2}» 46 mmHg. At the other extreme, when an alveolar-capillary unit has a high / ratio, it behaves almost like dead space, and the small amount of blood leaving the unit has partial pressures of O_2 and CO_2 (P_{O_2}» 150 mmHg, P_{CO_2}» 0 mmHg while breathing room air) approaching the composition of inspired gas.

In the ideal situation, all alveolar-capillary units have equal matching of ventilation and perfusion, i.e., a ratio of approximately 1 when each is expressed in L/min. However, even in the normal individual, some / mismatching is present, since there is normally a gradient of blood flow from the apices to the bases of the lungs. There is a similar gradient of ventilation from the apices to the bases, but it is less marked than the perfusion gradient. As a result, ventilation-perfusion ratios are higher at the lung apices than at the lung bases. Therefore, blood coming from the apices has a higher P_{O_2} and lower P_{CO_2} than blood coming from the bases. The net P_{O_2} and P_{CO_2} of the blood mixture coming from all areas of the lung is a flow-weighted average of the individual components, which reflects both the relative amount of blood from each unit and the O_2 and CO_2 *content* of the blood coming from each unit. Because of the sigmoid shape of the oxyhemoglobin dissociation curve (see Fig. 106-2), it is important to distinguish between the partial pressure and the content of O_2 in blood. Hemoglobin is almost fully

(~90%) saturated at a P_{O_2} of 60 mmHg, and little additional O_2 is carried by hemoglobin even with a substantial elevation of P_{O_2}above 60 mmHg. On the other hand, significant O_2desaturation of hemoglobin occurs once P_{O_2}falls below 60 mmHg and onto the steep descending limb of the curve. As a result, blood coming from regions of the lung with a high / ratio and a high P_{O_2} has only a small elevation in O_2content and cannot compensate for blood coming from regions with a low / ratio and a low P_{O_2}, which has a significantly decreased O_2content. Although / mismatching can influence P_{CO_2}, this effect is less marked and is often overcome by an increase in overall minute ventilation.

MEASUREMENT OF GAS EXCHANGE

Arterial Blood Gases The most commonly used measures of gas exchange are the partial pressures of O_2 and CO_2 in arterial blood, i.e., Pa_{O_2} and Pa_{CO_2}, respectively. These partial pressures do not measure directly the quantity of O_2 and CO_2 in blood but rather the driving pressure for the gas in blood. The actual quantity or content of a gas in blood also depends on the solubility of the gas in plasma and the ability of any component of blood to react with or bind the gas of interest. Since hemoglobin is capable of binding large amounts of O_2, oxygenated hemoglobin is the primary form in which O_2 is transported in blood. The actual content of O_2 in blood therefore depends both on the hemoglobin concentration and on the Pa_{O_2}. The Pa_{O_2}determines what percentage of hemoglobin is saturated with O_2, based on the position on the oxyhemoglobin dissociation curve. Oxygen content in normal blood (at 37°C, pH 7.4) can be determined by adding the amount of O_2dissolved in plasma to the amount bound to hemoglobin, according to the equation

since each gram of hemoglobin is capable of carrying 1.34 mL O_2 when fully saturated, and the amount of O_2 that can be dissolved in plasma is proportional to the P_{O_2}, with 0.0031 mL O_2dissolved per deciliter of blood per mmHg P_{O_2}. In arterial blood, the amount of O_2transported dissolved in plasma (approximately 0.3 mL O_2 per deciliter of blood) is trivial compared with the amount bound to hemoglobin (approximately 20 mL O_2 per deciliter of blood).

Most commonly, P_{O_2} is the measurement used to assess the effect of respiratory disease on the oxygenation of arterial blood. Direct measurement of O_2saturation in arterial blood by oximetry is also important in selected clinical conditions. For example, in patients with carbon monoxide intoxication, carbon monoxide preferentially displaces O_2 from hemoglobin, essentially making a portion of hemoglobin unavailable for binding to O_2. In this circumstance, carbon monoxide saturation is high and O_2saturation is low, even though the driving pressure for O_2 to bind to hemoglobin, reflected by P_{O_2}, is normal. Measurement of O_2saturation is also important for the determination of O_2content when mixed venous blood is sampled from a pulmonary arterial catheter to calculate cardiac output by the Fick technique. In mixed venous blood, the P_{O_2} is normally about 40 mmHg, but small changes in P_{O_2} may reflect relatively large changes in O_2saturation.

A useful calculation in the assessment of oxygenation is the alveolar-arterial O_2difference (PA_{O_2}-Pa_{O_2}), commonly called the *alveolar-arterial O_2gradient* (or A - a

gradient). This calculation takes into account the fact that alveolar and, hence, arterial P_{O_2} can be expected to change depending on the level of alveolar ventilation, reflected by the arterial P_{CO_2}. When a patient hyperventilates and has a low P_{CO_2} in arterial blood and alveolar gas, alveolar and arterial P_{O_2} will rise; conversely, hypoventilation and a high P_{CO_2} are accompanied by a decrease in alveolar and arterial P_{O_2}. These changes in arterial P_{O_2} are independent of abnormalities in O_2transfer at the alveolar-capillary level and reflect only the dependence of alveolar P_{O_2} on the level of alveolar ventilation.

In order to determine the alveolar-arterial O_2difference, the alveolar P_{O_2}(PA_{O_2}) must first be calculated. The equation most commonly used for this purpose, a simplified form of the alveolar gas equation, is

where FI_{O_2}= fractional concentration of inspired O_2(»0.21 when breathing room air); P_B= barometric pressure (approximately 760 mmHg at sea level); P_{H_2O}= water vapor pressure (47 mmHg when air is fully saturated at 37°C); and R= respiratory quotient (the ratio of CO_2production to O_2consumption, usually assumed to be 0.8). If the preceding values are substituted into the equation for the patient breathing air at sea level, the equation becomes

The alveolar-arterial O_2difference can then be calculated by subtracting measured $Pa_{O_2}$from calculated PA_{O_2}. In a healthy young person breathing room air, the PA_{O_2}-Pa_{O_2} is normally less than 15 mmHg; this value increases with age and may be as high as 30 mmHg in elderly patients.

The adequacy of CO_2elimination is measured by the partial pressure of CO_2 in arterial blood, i.e., Pa_{CO_2}. A more complete understanding of the mechanisms and chronicity of abnormal levels of P_{CO_2}also requires measurement of pH and/or bicarbonate (HCO_3^-), since P_{CO_2} and the patient's acid-base status are so closely intertwined (Chap. 50).

Pulse Oximetry Because measurement of Pa_{O_2}requires arterial puncture, it is not ideal either for office use or for routine or frequent measurement in the inpatient setting. Additionally, because it provides intermittent rather than continuous data about the patient's oxygenation, it is not ideal for close monitoring of unstable patients. Pulse oximetry, an alternative method for assessing oxygenation, is readily available in many clinical settings. Using a probe usually clipped over a patient's finger, the pulse oximeter calculates oxygen saturation (rather than Pa_{O_2}) based on measurements of absorption of two wavelengths of light by hemoglobin in pulsatile, cutaneous arterial blood. Because of differential absorption of the two wavelengths of light by oxygenated and nonoxygenated hemoglobin, the percentage of hemoglobin that is saturated with oxygen, i.e., the Sa_{O_2}, can be calculated and displayed instantaneously.

Although the pulse oximeter has been a major advance in the noninvasive, continuous monitoring of oxygenation, there are several issues and potential problems concerning its use. First, the clinician must be aware of the relationship between oxygen saturation and tension as shown by the oxyhemoglobin dissociation curve (Fig. 106-2). Because

the curve becomes relatively flat above an arterial P_{O_2} of 60 mmHg (corresponding to Sa_{O_2}= 90%), the oximeter is relatively insensitive to changes in Pa_{O_2}above this level. In addition, the position of the curve and therefore the specific relationship between Pa_{O_2} and Sa_{O_2} may change depending on factors such as temperature, pH, and the erythrocyte concentration of 2,3-diphosphoglycerate. Second, when cutaneous perfusion is decreased, e.g., owing to low cardiac output or the use of vasoconstrictors, the signal from the oximeter may be less reliable or even unobtainable. Third, other forms of hemoglobin, such as carboxyhemoglobin and methemoglobin, are not distinguishable from oxyhemoglobin when only two wavelengths of light are used. The Sa_{O_2}values reported by the pulse oximeter are not reliable in the presence of significant amounts of either of these forms of hemoglobin. In contrast, the device used to measure oxygen saturation in samples of arterial blood, called the CO-oximeter, uses at least four wavelengths of light and is capable of distinguishing oxyhemoglobin, deoxygenated hemoglobin, carboxyhemoglobin, and methemoglobin. Finally, the clinician must remember that the often-used goal of Sa_{O_2}[3] 90% does not indicate anything about CO_2elimination and therefore does not ensure a clinically acceptable P_{CO_2}.

Diffusing Capacity The ability of gas to diffuse across the alveolar-capillary membrane is ordinarily assessed by the diffusing capacity of the lung for carbon monoxide (DL_{CO}). In this test, a small concentration of carbon monoxide (0.3%) is inhaled, usually in a single breath that is held for approximately 10 s. The carbon monoxide is diluted by the gas already present in the alveoli and is also taken up by hemoglobin as the erythrocytes course through the pulmonary capillary system. The concentration of carbon monoxide in exhaled gas is measured, and DL_{CO} is calculated as the quantity of carbon monoxide absorbed per minute per mmHg pressure gradient from the alveoli to the pulmonary capillaries. The value obtained for DL_{CO}depends on the alveolar-capillary surface area available for gas exchange and on the pulmonary capillary blood volume. In addition, the thickness of the alveolar-capillary membrane, the degree of / mismatching, and the patient's hemoglobin level will affect the measurement. Because of this effect of hemoglobin levels on DL_{CO}, the measured DL_{CO} is frequently corrected to take the patient's hemoglobin level into account. The value for DL_{CO}, ideally corrected for hemoglobin, can then be compared with a predicted value, based either on age, height, and gender or on the alveolar volume (VA) at which the value was obtained. Alternatively, the DL_{CO} can be divided by VA and the resulting value for DL_{CO}/VA compared with a predicted value.

Approach to the Patient

Arterial Blood Gases Hypoxemia is a common manifestation of a variety of diseases affecting the lungs or other parts of the respiratory system. The broad clinical problem of hypoxemia is often best characterized according to the underlying mechanism. The four basic, and not mutually exclusive, mechanisms of hypoxemia are (1) a decrease in inspired P_{O_2}, (2) hypoventilation, (3) shunting, and (4) / mismatching. Hypoxemia due to decreased diffusion occurs only under selected clinical circumstances and is not usually included among the general categories of hypoxemia. Determining the underlying mechanism for hypoxemia depends on measurement of the Pa_{CO_2}, calculation of PA_{O_2}-Pa_{O_2}, and knowledge of the response to supplemental O_2. A flowchart summarizing the approach to the hypoxemic patient is given inFig. 250-5.

A decrease in the inspired P_{O_2} and hypoventilation both cause hypoxemia by lowering $P_{A_{O_2}}$ and therefore $P_{a_{O_2}}$. In each case, gas exchange at the alveolar-capillary level occurs normally, and $P_{A_{O_2}}$-$P_{a_{O_2}}$ is not elevated. Hypoxemia due to decreased inspired P_{O_2} can be diagnosed from knowledge of the clinical situation. Inspired P_{O_2} is lowered either because the patient is at a high altitude, where barometric pressure is low, or, much less commonly, because the patient is breathing a gas mixture containing less than 21% O_2. The hallmark of hypoventilation as a cause of hypoxemia is an elevation in $P_{a_{CO_2}}$. This is associated with an increase in $P_{A_{CO_2}}$and a fall in $P_{A_{O_2}}$. When hypoxemia is due purely to a low inspired P_{O_2} or to alveolar hypoventilation, $P_{A_{O_2}}$-$P_{a_{O_2}}$ is normal. If $P_{A_{O_2}}$-$P_{a_{O_2}}$ and $P_{a_{CO_2}}$are both elevated, then an additional mechanism, such as / mismatching or shunting, is contributing to hypoxemia.

Shunting is a cause of hypoxemia when desaturated blood effectively bypasses oxygenation at the alveolar-capillary level. This situation occurs either because a structural problem allows desaturated blood to bypass the normal site of gas exchange or because perfused alveoli are not ventilated. Shunting is associated with an elevation in the $P_{A_{O_2}}$-$P_{a_{O_2}}$value. When shunting is an important contributing factor to hypoxemia, the lowered $P_{a_{O_2}}$ is relatively refractory to improvement by supplemental O_2.

Finally, the largest clinical category of hypoxemia is / mismatching. With / mismatching, regions with low / ratios contribute blood with a low P_{O_2} and a low O_2content. Corresponding regions with high / ratios contribute blood with a high P_{O_2}. However, because blood is already almost fully saturated at a normal P_{O_2}, elevation of the P_{O_2} to a high value does not significantly increase O_2saturation or content and therefore cannot compensate for the reduction of O_2saturation and content in blood coming from regions with a low / ratio. When / mismatch is the primary cause of hypoxemia, $P_{A_{O_2}}$-$P_{a_{O_2}}$ is elevated, and P_{CO_2}generally is normal. Supplemental O_2corrects the hypoxemia by raising the P_{O_2} in blood coming from regions with a low / ratio; this response distinguishes hypoxemia due to / mismatching from that due to true shunt.

The essential mechanism underlying all cases of hypercapnia is alveolar ventilation that is inadequate for the amount of CO_2produced. It is conceptually useful to characterize CO_2retention further, based on a more detailed examination of the potential contributing factors. These include (1) increased CO_2production; (2) decreased ventilatory drive ("won't breathe"); (3) malfunction of the respiratory pump or increased airways resistance, which makes it more difficult to sustain adequate ventilation ("can't breathe"); and (4) inefficiency of gas exchange (increased dead space or / mismatch) necessitating a compensatory increase in overall minute ventilation. In practice, more than one of these mechanisms is commonly responsible for hypercapnia, since increased minute ventilation is capable of compensating for increased CO_2production and for inefficiencies of gas exchange.

Diffusing Capacity Although abnormalities in diffusion are rarely responsible for hypoxemia, clinical measurement of diffusing capacity is frequently used to assess the functional integrity of the alveolar-capillary membrane, which includes the pulmonary capillary bed. Diseases that affect solely the airways generally do not lower DL_{CO}, whereas diseases that affect the alveolar walls or the pulmonary capillary bed will have an effect on DL_{CO}. Even though DL_{CO} is a useful marker for assessing whether disease affecting the alveolar-capillary bed is present, an abnormal DL_{CO}does not necessarily

imply that diffusion limitation is responsible for hypoxemia in a particular patient.

CLINICAL CORRELATIONS

Useful clinical correlations can be made with the mechanisms underlying hypoxemia (Fig. 250-5). A lowered inspired P_{O_2} contributes to hypoxemia if either the patient is at high altitude or if the concentration of inspired O_2 is less than 21%. The latter problem occurs if a patient receiving anesthesia or ventilatory support is inadvertently given a gas mixture to breathe containing less than 21% O_2 or if O_2 is consumed from the ambient gas, as can occur during smoke inhalation from a fire. The primary feature of hypoventilation as a cause of hypoxemia is an elevation in arterial P_{CO_2}. *For further discussion of the clinical correlations with hypoventilation, see Chap. 263.*

Shunting as a cause of hypoxemia can reflect transfer of blood from the right to the left side of the heart without passage through the pulmonary circulation, as occurs with an intracardiac shunt. This problem is most common in the setting of cyanotic congenital heart disease, when an interatrial or interventricular septal defect is associated with pulmonary hypertension so that shunting is in the right-to-left rather than the left-to-right direction. Shunting of blood through the pulmonary parenchyma is most frequently due to disease causing absence of ventilation to perfused alveoli. This can occur if the alveoli are atelectatic or if they are filled with fluid, as in pulmonary edema (both cardiogenic and noncardiogenic), or with extensive intraalveolar exudation of fluid due to pneumonia. Less commonly, vascular anomalies with arteriovenous shunting in the lung can cause hypoxemia. These anomalies can be hereditary, as found with hereditary hemorrhagic telangiectasia (Osler-Rendu-Weber syndrome), or acquired, as in pulmonary vascular malformations secondary to hepatic cirrhosis, which are similar to the commonly recognized cutaneous vascular malformations ("spider hemangiomas").

Ventilation-perfusion mismatch is the most common cause of hypoxemia clinically. Most of the processes affecting either the airways or the pulmonary parenchyma are distributed unevenly throughout the lungs and do not necessarily affect ventilation and perfusion equally. Some areas of lung may have good perfusion and poor ventilation, whereas others may have poor perfusion and relatively good ventilation. Important examples of airways diseases in which / mismatch causes hypoxemia are asthma and chronic obstructive lung disease. Parenchymal lung diseases causing / mismatch and hypoxemia include interstitial lung disease and pneumonia.

Clinically important alterations in CO_2 elimination range from excessive ventilation and hypocapnia to inadequate CO_2 elimination and hypercapnia. *For further discussion of these clinical problems, see Chap. 263.*

Diffusing Capacity Measurement of DL_{CO} may be useful for assessing disease affecting the alveolar-capillary bed or the pulmonary vasculature. In practice, three main categories of disease are associated with lowered DL_{CO}: interstitial lung disease, emphysema, and pulmonary vascular disease. With interstitial lung disease, scarring of alveolar-capillary units diminishes the area of the alveolar-capillary bed as well as pulmonary blood volume. With emphysema, alveolar walls are destroyed, so the surface area of the alveolar-capillary bed is again diminished. In patients with disease causing a decrease in the cross-sectional area and volume of the pulmonary vascular bed, such

as recurrent pulmonary emboli or primary pulmonary hypertension, DL_{CO} is commonly diminished.

Diffusing capacity may be elevated if pulmonary blood volume is increased, as may be seen in congestive heart failure. However, once interstitial and alveolar edema ensue, the net DL_{CO}depends on the opposing influences of increased pulmonary capillary blood volume elevating DL_{CO} and pulmonary edema decreasing it. Finding an elevated DL_{CO} may be useful in the diagnosis of alveolar hemorrhage, as in Goodpasture's syndrome. Hemoglobin contained in erythrocytes in the alveolar lumen is capable of binding carbon monoxide, so the exhaled carbon monoxide concentration is diminished and the measured DL_{CO} is increased.

(Bibliography omitted in Palm version)

Back to Table of Contents

251. DIAGNOSTIC PROCEDURES IN RESPIRATORY DISEASE - *Steven E. Weinberger*, *Jeffrey M. Drazen*

The diagnostic modalities available for assessing the patient with suspected or known respiratory system disease include imaging studies and techniques for acquiring biologic specimens, some of which involve direct visualization of part of the respiratory system.*Methods used to characterize the functional changes developing as a result of disease, including pulmonary function tests and measurements of gas exchange, are discussed in Chap. 250.*

IMAGING STUDIES

ROUTINE RADIOGRAPHY

Routine chest radiography, which generally includes both posteroanterior and lateral views, is an integral part of the diagnostic evaluation of diseases involving the pulmonary parenchyma, the pleura, and, to a lesser extent, the airways and the mediastinum (see Figs. 249-1 and249-2). Lateral decubitus views are often useful for determining whether pleural abnormalities represent freely flowing fluid, whereas apical lordotic views can often visualize disease at the lung apices better than the standard posteroanterior view. Portable equipment, which is often used for acutely ill patients who either cannot be transported to a radiology suite or cannot stand up for posteroanterior and lateral views, generally yields just a single radiograph taken in the anteroposterior direction.**Common radiographic patterns and their clinical correlates are reviewed in Chap. 249.*

COMPUTED TOMOGRAPHY

Computed tomography (CT) offers several advantages over routine chest radiography. First, the use of cross-sectional images often makes it possible to distinguish between densities that would be superimposed on plain radiographs. Second, CT is far better than routine radiographic studies at characterizing tissue density, distinguishing subtle differences in density between adjacent structures, and providing accurate size assessment of lesions. As a result, CT is particularly valuable in assessing hilar and mediastinal disease (which is often poorly characterized by plain radiography), in identifying and characterizing disease adjacent to the chest wall or spine (including pleural disease), and in identifying areas of fat density or calcification in pulmonary nodules (Fig. 251-1). Its utility in the assessment of mediastinal disease has made CT an important tool in the staging of lung cancer (Chap. 88), as an assessment of tumor involvement of mediastinal lymph nodes is critical to proper staging. With the additional use of contrast material, CT also makes it possible to distinguish vascular from nonvascular structures, which is particularly important in distinguishing lymph nodes and masses from vascular structures.

HelicalCTscanning allows the collection of continuous data over a larger volume of lung during a single breath-holding maneuver than is possible with conventional CT. With CT angiography, in which intravenous contrast is administered and images are acquired rapidly by helical scanning, pulmonary emboli can be detected in segmental and larger pulmonary arteries. With high-resolution CT (HRCT), the thickness of individual

cross-sectional images is approximately 1 to 2 mm, rather than the usual 10 mm, and the images are reconstructed with high-spatial-resolution algorithms. The detail that can be seen on HRCT scans allows better recognition of subtle parenchymal and airway disease, such as bronchiectasis, emphysema, and diffuse parenchymal disease (Fig. 251-2). Certain nearly pathognomonic patterns have now been recognized for many of the interstitial lung diseases, such as lymphangitic carcinoma, idiopathic pulmonary fibrosis, sarcoidosis, and eosinophilic granuloma; at present it is not yet clear in what settings these patterns will obviate the need for obtaining lung tissue.

MAGNETIC RESONANCE IMAGING

The role of magnetic resonance imaging (MRI) in the evaluation of respiratory system disease is less well defined than that ofCT. Because MRI generally provides a less detailed view of the pulmonary parenchyma as well as poorer spatial resolution, its usefulness in the evaluation of parenchymal lung disease is limited at present. However, MRI has advantages over CT in certain clinical settings. Because its images can be reconstructed in sagittal and coronal as well as transverse planes, MRI may be better for imaging abnormalities near the lung apex, the spine, and the thoracoabdominal junction. In addition, vascular structures can be distinguished from nonvascular structures without the need for contrast. Flowing blood does not produce a signal on MRI, so vessels appear as hollow tubular structures. This feature can be useful in determining whether abnormal hilar or mediastinal densities are vascular in origin and in defining aortic lesions such as aneurysms or dissection.

SCINTIGRAPHIC IMAGING

Radioactive isotopes, administered by either intravenous or inhaled routes, allow the lungs to be imaged with a gamma camera. The most common use of such imaging is ventilation-perfusion lung scanning performed for evaluation of pulmonary embolism. When injected intravenously, albumin macroaggregates labeled with technetium 99m become lodged in pulmonary capillaries; therefore, the distribution of the trapped radioisotope follows the distribution of blood flow. When inhaled, radiolabeled xenon gas can be used to demonstrate the distribution of ventilation. For example, pulmonary thromboembolism usually produces one or more regions of ventilation-perfusion mismatch -- that is, regions in which there is a defect in perfusion that follows the distribution of a vessel and that is not accompanied by a corresponding defect in ventilation (Chap. 261). Another common use of such radioisotope scans is in a patient with impaired lung function who is being considered for lung resection. The distribution of the isotope(s) can be used to assess the regional distribution of blood flow and ventilation, allowing the physician to estimate the level of postoperative lung function.

Another scintigraphic imaging technique, gallium imaging, has been of diagnostic value in patients with *Pneumocystis carinii* pneumonia and other opportunistic infections. Use of gallium imaging may provide clues to sort out the differential diagnosis of pulmonary infiltrates in immunosuppressed patients, especially patients with AIDS.

PULMONARY ANGIOGRAPHY

The pulmonary arterial system can be visualized by pulmonary angiography, in which

radiopaque contrast medium is injected through a catheter previously threaded into the pulmonary artery. When performed in cases of pulmonary embolism, pulmonary angiography demonstrates the consequences of an intravascular clot -- either a defect in the lumen of a vessel (a "filling defect") or an abrupt termination ("cutoff") of the vessel. Other, less common indications for pulmonary angiography include visualization of a suspected pulmonary arteriovenous malformation and assessment of pulmonary arterial invasion by a neoplasm.

ULTRASOUND

Because ultrasound energy is rapidly dissipated in air, ultrasound imaging is not useful for evaluation of the pulmonary parenchyma. However, it is helpful in the detection and localization of pleural abnormalities and is often used as a guide to placement of a needle for sampling of pleural liquid (i.e., for thoracentesis).

TECHNIQUES FOR OBTAINING BIOLOGIC SPECIMENS

COLLECTION OF SPUTUM

Sputum can be collected either by spontaneous expectoration or after inhalation of an irritating aerosol, such as hypertonic saline. The latter method, called *sputum induction*, is commonly used to obtain sputum for diagnostic studies, either because sputum is not spontaneously being produced or because of an expected higher yield of certain types of findings. Knowledge of the appearance and quality of the sputum specimen obtained is especially important when one is interested in Gram's staining and culture. Because sputum consists mainly of secretions from the tracheobronchial tree rather than the upper airway, the finding of alveolar macrophages and other inflammatory cells is consistent with a lower respiratory tract origin of the sample, whereas the presence of squamous epithelial cells in a "sputum" sample indicates contamination by secretions from the upper airways.

Besides processing for routine bacterial pathogens by Gram's staining and culture, sputum can be processed for a variety of other pathogens, including staining and culture for mycobacteria or fungi, culture for viruses, and staining for *P. carinii*. In the specific case of sputum obtained for evaluation of *P. carinii* pneumonia in a patient infected with HIV, for example, sputum should be collected by induction, rather than spontaneous expectoration, and an immunofluorescent stain should be used to detect the organisms. Cytologic staining of sputum for malignant cells, using the traditional Papanicolaou method, allows noninvasive evaluation for suspected lung cancer. Traditional stains and cultures are now also being supplemented in some cases by immunologic techniques and by molecular biologic methods, including the use of polymerase chain reaction amplification and DNA probes.

PERCUTANEOUS NEEDLE ASPIRATION

A needle can be inserted through the chest wall into a pulmonary lesion for the purpose of aspirating material for analysis by cytologic or microbiologic techniques. The procedure is usually carried out underCTguidance, which assists in the positioning of the needle and assures that it is localized in the lesion. Although the potential risks of

this procedure include intrapulmonary bleeding and creation of a pneumothorax with collapse of the underlying lung, the low risk of complication in experienced hands is usually worth the information obtained. However, a limitation of the technique is sampling error due to the small amount of material obtained. Thus, findings other than a specific cytologic or microbiologic diagnosis are of limited clinical value.

THORACENTESIS

Sampling of pleural liquid by thoracentesis is commonly performed for diagnostic purposes or, in the case of a large effusion, for palliation of dyspnea. Diagnostic sampling, either by blind needle aspiration or after localization by ultrasound, allows the collection of liquid for microbiologic and cytologic studies. Analysis of the fluid obtained for its cellular composition and chemical constituents, including glucose, protein, and lactate dehydrogenase, allows the effusion to be classified as either exudative or transudative (Chap. 262). In some cases, particularly in the setting of possible tuberculous involvement of the pleura (tuberculous pleuritis), closed biopsy of the parietal pleura is also performed, using a cutting needle (either an Abrams or a Cope biopsy needle) to sample tissue for histopathologic examination and culture.

BRONCHOSCOPY

Bronchoscopy is the process of direct visualization of the tracheobronchial tree. Bronchoscopy with a rigid bronchoscope is generally performed in an operating room on a patient under general anesthesia. The development of a flexible fiberoptic bronchoscope has revolutionized the diagnostic use of bronchoscopy. Although bronchoscopy is now performed almost exclusively with fiberoptic instruments, rigid bronchoscopes still have a role in selected circumstances, primarily because of their larger suction channel and the fact that the patient can be ventilated through the bronchoscope channel. These situations include the retrieval of a foreign body and the suctioning of a massive hemorrhage, for which the small suction channel of the bronchoscope may be insufficient.

Flexible Fiberoptic Bronchoscopy This is an outpatient procedure that is usually performed in an awake but sedated patient. The bronchoscope is passed through either the mouth or the nose, between the vocal cords, and into the trachea. The ability to flex the scope makes it possible to visualize virtually all airways to the level of subsegmental bronchi. The bronchoscopist is able to identify endobronchial pathology, including tumors, granulomas, bronchitis, foreign bodies, and sites of bleeding. Samples from airway lesions can be taken by several methods, including washing, brushing, and biopsy. Washing involves instillation of sterile saline through a channel of the bronchoscope and onto the surface of a lesion. A portion of the liquid is collected by suctioning through the bronchoscope, and the recovered material can be analyzed for cells (cytology) or organisms (by standard stains and cultures). Brushing or biopsy of the surface of the lesion, using a small brush or biopsy forceps at the end of a long cable inserted through a channel of the bronchoscope, allows recovery of cellular material or tissue for analysis by standard cytologic and histopathologic methods.

The bronchoscope can be used to sample material not only from the regions that can be directly visualized (i.e., the airways) but also from the more distal pulmonary

parenchyma. With the bronchoscope wedged into a subsegmental airway, aliquots of sterile saline can be instilled through the scope, allowing sampling of cells and organisms even from alveolar spaces. This procedure, called *bronchoalveolar lavage*, has been particularly useful for the recovery of organisms such as *P. carinii* in patients with HIV infection.

Brushing and biopsy of the distal lung parenchyma can also be performed with the same instruments that are used for endobronchial sampling. These instruments can be passed through the scope into small airways, where they penetrate the airway wall, allowing biopsy of peribronchial alveolar tissue. This procedure, called *transbronchial biopsy*, is used when there is either relatively diffuse disease or a localized lesion of adequate size. With the aid of fluoroscopic imaging, the bronchoscopist is able to determine not only whether and when the instrument is in the area of abnormality, but also the proximity of the instrument to the pleural surface. If the forceps are too close to the pleural surface, there is a risk of violating the visceral pleura and creating a pneumothorax; the other potential complication of transbronchial biopsy is pulmonary hemorrhage. The incidence of these complications is less than several percent.

Another procedure involves use of a hollow-bore needle passed through the bronchoscope for sampling of tissue adjacent to the trachea or a large bronchus. The needle is passed through the airway wall, and cellular material can be aspirated from mass lesions or enlarged lymph nodes, generally in a search for malignant cells. This procedure can facilitate the staging of lung cancer by identifying mediastinal lymph node involvement and in some cases obviates the need for a more invasive procedure.

The bronchoscope may provide the opportunity for treatment as well as diagnosis. For example, an aspirated foreign body may be retrieved with an instrument passed through the scope, and bleeding may be controlled with a balloon catheter similarly introduced. Newer interventional techniques performed through a bronchoscope include methods for achieving and maintaining patency of airways that are partially or completely occluded, especially by tumors. These techniques include laser therapy, cryotherapy, electrocautery, and stent placement.

VIDEO-ASSISTED THORACIC SURGERY

Recent advances in video technology have allowed the development of thoracoscopy, or video-assisted thoracic surgery (VATS), for the diagnosis and management of pleural as well as parenchymal lung disease. This procedure, done under general anesthesia, involves the passage of a rigid scope with a distal lens through a trocar inserted into the pleura. A high-quality image is shown on a monitor screen, allowing the operator to manipulate instruments passed into the pleural space through separate small intercostal incisions. With these instruments, the operator can biopsy lesions of the pleura under direct vision, which provides an obvious advantage over closed pleural biopsy. In addition, this procedure is now used commonly to biopsy peripheral lung tissue or to remove peripheral nodules, for both diagnostic and therapeutic purposes. Because this procedure is much less invasive than the traditional thoracotomy performed for lung biopsy, it has largely supplanted "open lung biopsy."

THORACOTOMY

Although frequently replaced byVATS, thoracotomy remains an option for the diagnostic sampling of lung tissue. It provides the largest amount of material, and it can be used to biopsy and/or excise lesions that are too deep or too close to vital structures for removal by VATS. The choice between VATS and thoracotomy needs to be made on a case-by-case basis, and the relative indications for each are still evolving as more experience is being gained with VATS.

MEDIASTINOSCOPY AND MEDIASTINOTOMY

Tissue biopsy is often critical for the diagnosis of mediastinal masses or enlarged mediastinal lymph nodes. AlthoughCT is useful for determining the size of mediastinal lymph nodes as part of the staging of lung cancer, confirmation that enlarged lymph nodes are actually involved with tumor generally requires biopsy and histopathologic examination. The two major procedures used to obtain specimens from masses or nodes in the mediastinum are mediastinoscopy (via a suprasternal approach) and mediastinotomy (via a parasternal approach). Both procedures are performed under general anesthesia by a qualified surgeon. In the case of suprasternal mediastinoscopy, a rigid mediastinoscope is inserted at the suprasternal notch and passed into the mediastinum along a pathway just anterior to the trachea. Tissue can be obtained with biopsy forceps passed through the scope, sampling masses or nodes that are in a paratracheal or pretracheal position. Left paratracheal and aortopulmonary lymph nodes are not accessible by this route and thus are commonly sampled by parasternal mediastinotomy (the Chamberlain procedure). This approach involves either a right or left parasternal incision and dissection directly down to a mass or node that requires biopsy.

(Bibliography omitted in Palm version)

Back to Table of Contents

SECTION 2 -DISEASES OF THE RESPIRATORY SYSTEM

252. ASTHMA - *E. R. McFadden, Jr.*

DEFINITION

Asthma is defined as a chronic inflammatory disease of airways that is characterized by increased responsiveness of the tracheobronchial tree to a multiplicity of stimuli. It is manifested physiologically by a widespread narrowing of the air passages, which may be relieved spontaneously or as a result of therapy, and clinically by paroxysms of dyspnea, cough, and wheezing. Asthma is an episodic disease, with acute exacerbations interspersed with symptom-free periods. Typically, most attacks are short-lived, lasting minutes to hours, and clinically the patient seems to recover completely after an attack. However, there can be a phase in which the patient experiences some degree of airway obstruction daily. This phase can be mild, with or without superimposed severe episodes, or much more serious, with severe obstruction persisting for days or weeks; the latter condition is known as *status asthmaticus*. In unusual circumstances, acute episodes can cause death.

PREVALENCE AND ETIOLOGY

Asthma is very common; it is estimated that 4 to 5% of the population of the United States is affected. Similar figures have been reported from other countries. Bronchial asthma occurs at all ages but predominantly in early life. About one-half of cases develop before age 10, and another third occur before age 40. In childhood, there is a 2:1 male/female preponderance, but the sex ratio equalizes by age 30.

From an etiologic standpoint, asthma is a heterogeneous disease. It is useful for epidemiologic and clinical purposes to classify asthma by the principal stimuli that incite or are associated with acute episodes. However, it is important to emphasize that this distinction may often be artificial, and the response of a given subclassification usually can be initiated by more than one type of stimulus. Furthermore, the application of molecular and cell biologic techniques to asthma pathogenesis is also beginning to blur this type of classification. With these reservations in mind, one can describe two broad types of asthma: allergic and idiosyncratic.

Atopy is the single largest risk factor for the development of asthma. *Allergic asthma* is often associated with a personal and/or family history of allergic diseases such as rhinitis, urticaria, and eczema, with positive wheal-and-flare skin reactions to intradermal injection of extracts of airborne antigens, with increased levels of IgE in the serum, and/or with a positive response to provocation tests involving the inhalation of specific antigen.

A significant fraction of patients with asthma present with no personal or family history of allergy, with negative skin tests, and with normal serum levels of IgE, and therefore have disease that cannot be classified on the basis of defined immunologic mechanisms. These patients are said to have *idiosyncratic asthma*. Many develop a typical symptom complex on contracting an upper respiratory illness. The initial insult may be little more than a common cold, but after several days the patient begins to

develop paroxysms of wheezing and dyspnea that can last for days to months. These individuals should not be confused with persons in whom the symptoms of bronchospasm are superimposed on chronic bronchitis or bronchiectasis (Chaps. 256 and258).

Many patients have disease that does not fit clearly into either of the preceding categories but instead falls into a mixed group with features of each. In general, asthma that has its onset in early life tends to have a strong allergic component, whereas asthma that develops late tends to be nonallergic or to have a mixed etiology.

PATHOGENESIS OF ASTHMA

The common denominator underlying the asthmatic diathesis is a nonspecific hyperirritability of the tracheobronchial tree. When airway reactivity is high, symptoms are more severe and persistent, and the amount of therapy required to control the patient's complaints is greater. In addition, the magnitude of diurnal fluctuations in lung function is greater, and the patient tends to awaken at night or in the early morning with breathlessness.

In both normal and asthmatic individuals, airway reactivity rises after viral infections of the respiratory tract and exposure to oxidant air pollutants such as ozone and nitrogen dioxide (but not sulfur dioxide). Viral infections have more profound consequences, and airway responsiveness may remain elevated for many weeks after a seemingly trivial upper respiratory tract infection. In contrast, airway reactivity remains high for only a few days after exposure to ozone. Allergens can cause airway responsiveness to rise within minutes and to remain elevated for weeks. If the dose of antigen is high enough, acute episodes of obstruction may occur daily for a prolonged period after a single exposure.

The most popular hypothesis at present for the pathogenesis of asthma is that it derives from a state of persistent subacute inflammation of the airways. An active inflammatory process is frequently observed in endobronchial biopsy specimens even from asymptomatic patients. The airways can be edematous and infiltrated with eosinophils, neutrophils, and lymphocytes, with or without an increase in the collagen content of the epithelial basement membrane. There may also be glandular hypertrophy. The most ubiquitous finding is a generalized increase in cellularity associated with an elevated capillary density. Occasionally, denudation of the epithelium may also be observed.

Although the translation of these histologic observations into a disease process is still incomplete, it is widely believed that the physiologic and clinical features of asthma derive from an interaction among the resident and infiltrating inflammatory cells in the airway surface epithelium, inflammatory mediators, and cytokines. The cells thought to play important parts in the inflammatory response are mast cells, eosinophils, lymphocytes, and epithelial cells. The roles of neutrophils and macrophages are less well defined. Each of these cell types can contribute mediators and cytokines to initiate and amplify both acute inflammation and the long-term pathologic changes described above. The mediators released -- histamine; bradykinin; the leukotrienes C, D, and E; platelet-activating factor; and prostaglandins (PGs) E_2, F_2a, and D_2 -- produce an intense, immediate inflammatory reaction involving bronchoconstriction, vascular congestion, and edema formation. In addition to their ability to evoke prolonged

contraction of airway smooth muscle and mucosal edema, the leukotrienes may also account for some of the other pathophysiologic features of asthma, such as increased mucus production and impaired mucociliary transport. This intense local event can then be followed by a more chronic one. The chemotactic factors elaborated (eosinophil and neutrophil chemotactic factors of anaphylaxis and leukotriene B_4) bring eosinophils, platelets, and polymorphonuclear leukocytes to the site of the reaction. These infiltrating cells as well as resident macrophages and the airway epithelium itself potentially are an additional source of mediators to enhance both the immediate and the cellular phase. The airway epithelium is both the target of, and a contributor to, the inflammatory cascade. These cells amplify bronchoconstriction by elaborating endothelin-1 and promoting vasodilatation through the release of nitric oxide,PG$E_2$and the 15-hydroxyeicosatetraenoic acid (15-HETE) products of arachidonic acid metabolism. They also generate cytokines such as granulocyte-macrophage colony stimulating factor (GM-CSF), interleukin (IL)8, Rantes, and eotaxin.

Like the mast cell in the early reaction, the eosinophil appears to play an important part in the infiltrative component. The granular proteins in this cell (major basic protein and eosinophilic cationic protein) and oxygen-derived free radical are capable of destroying the airway epithelium, which then is sloughed into the bronchial lumen in the form of Creola bodies. Besides resulting in a loss of barrier and secretory function, such damage elicits the production of chemotactic cytokines, leading to further inflammation. In theory, it also can expose sensory nerve endings, thus initiating neurogenic inflammatory pathways. That, in turn, could convert a primary local event into a generalized reaction via a reflex mechanism.

T lymphocytes also appear to be important in the inflammatory response. These cells are present in increased numbers in asthmatic airways and produce cytokines that activate cell-mediated immunity, as well as humoral (IgE) immune responses. Activated T cells recovered from the lungs of persons with asthma express messenger RNA for the cytokines known to play a part in the recruitment and activation of mast cells and eosinophils. Furthermore, the T_H1 and T_H2 lymphocyte subtypes have functions that may influence the asthmatic response. The T_H1 cytokinesIL-2 and interferon (IFN) g can promote the growth and differentiation of B cells and the activation of macrophages, respectively. The T_H2 cytokines IL-4 and IL-5 stimulate B-cell growth and immunoglobulin secretion, and IL-5 promotes eosinophil proliferation, differentiation, and activation. It can also facilitate granule release from basophils.

Cytokine production is another central component of the inflammation of asthma. Cytokines are synthesized and released from many of the inflammatory cells mentioned above, as well as from epithelial cells, fibroblasts, endothelial cells, and airway smooth muscle. Cytokines activate specific cell-surface receptors that are coupled to signal transduction pathways, which often result in alterations of gene regulation and enzyme production. The cytokines that are particularly relevant to asthma are secreted by T lymphocytes and includeIL-3 enhanced (mast cell survival), IL-4 and IL-13 (switching of B lymphocytes to IgE production and expression of adhesion molecules), and IL-5 (differentiation and enhanced survival of eosinophils). Other cytokines, such as IL-1B, IL-6, IL-11, tumor necrosis factor a(TNF-a) andGM-CSF, are proinflammatory and may amplify the inflammatory response.

The relative roles of each of the above elements in the production of heightened airway reactivity and clinical asthma have yet to be determined. Although inflammation is clearly important, recent evidence indicates that the intensity of the cellular infiltrate in the airways is not related either to the severity of the disease state or to the level of airway reactivity. Furthermore, it is unlikely that any one cell type or mediator accounts for every feature. For example, mast cell-derived mediators cannot explain the whole picture, for they have been found in the blood of individuals with mast cell-related diseases such as cold-induced and cholinergic-induced urticaria and in the airways of atopic individuals without asthma. Since these individuals had no lower respiratory illness or complaints, these alleged mediators of asthma would appear to need a unique background from which to exhibit their effects. Similarly, the inflammatory cells believed to be relevant to asthma are also found in the airways of atopic individuals without asthma, raising the possibility that they are merely nonspecific markers of atopy rather than specific indexes of asthma. Finally, the therapeutic administration ofIL-2 andGM-CSFto patients with cancer results in eosinophilia with cell activation but not in asthma.

GENETIC CONSIDERATIONS

Although there is little doubt that asthma has a strong familial component, the identification of the genetic mechanisms underlying the illness has proven difficult for multiple reasons, including such fundamental issues as a lack of uniform agreement on the definition of the disease, the inability to define a single phenotype, non-Mendelian modes of inheritance, and an incomplete understanding of how environmental factors modify genetic expression. Screening families for candidate genes has identified multiple chromosomal regions that relate to atopy, elevated IgE levels, and airway hyperresponsiveness. Evidence for genetic linkage of high total serum IgE levels and atopy has been observed on chromosomes 5q, 11q, and 12q in a number of populations scattered throughout the world. Regions of the genome demonstrating evidence for linkage to bronchial hyperreactivity also typically show evidence for linkage to elevated total serum IgE levels. Excellent candidate genes exist for specific abnormalities in asthma within the regions that were identified in the linkage studies. For example, chromosome 5q contains cytokine clusters includingIL-4, IL-5, IL-9, and IL-13. Other regions on chromosome 5q also contain the beta-adrenergic receptors and the glucocorticoid receptors. Chromosome 6p contains regions that are important in antigen presentation and mediation of the inflammatory response. Chromosome 12q contains two genes that could influence atopy and airway hyperresponsiveness, including nitric oxide synthase.

The stimuli that interact with airway responsiveness and incite acute episodes of asthma can be grouped into seven major categories: allergenic, pharmacologic, environmental, occupational, infectious, exercise-related, and emotional.

Allergens Allergic asthma is dependent on an IgE response controlled by T and B lymphocytes and activated by the interaction of antigen with mast cell-bound IgE molecules. The airway epithelium and submucosa contain dendritic cells that capture and process antigen. After taking up an immunogen, these cells migrate to the local lymph nodes where they present the material to T cell receptors. In the appropriate genetic setting, the interaction of antigen with a naive T cell T_H0 in the presence of IL-4

leads to the differentiation of the cell to a T_H2 subset. This process not only helps facilitate the inflammation of asthma but also causes B lymphocytes to switch their antibody production from IgG and IgM to IgE. Most of the allergens that provoke asthma are airborne, and to induce a state of sensitivity they must be reasonably abundant for considerable periods of time. Once sensitization has occurred, however, the patient can exhibit exquisite responsivity, so that minute amounts of the offending agent can produce significant exacerbations of the disease. Immune mechanisms appear to be causally related to the development of asthma in 25 to 35% of all cases and to be contributory in perhaps another third. Higher prevalences have been suggested, but it is difficult to know how to interpret the data because of confounding factors. Allergic asthma is frequently seasonal, and it is most often observed in children and young adults. A nonseasonal form may result from allergy to feathers, animal danders, dust mites, molds, and other antigens that are present continuously in the environment. Exposure to antigen typically produces an immediate response in which airway obstruction develops in minutes and then resolves. In 30 to 50% of patients, a second wave of bronchoconstriction, the so-called late reaction, develops 6 to 10 h later. In a minority, only a late reaction occurs. It was formerly thought that the late reaction was essential to the development of the increase in airway reactivity that follows antigen exposure. Recent data show that not to be the case.

The mechanism by which an inhaled allergen provokes an acute episode of asthma depends in part on antigen-antibody interactions on the surface of pulmonary mast cells, with the subsequent generation and release of the mediators of immediate hypersensitivity. Current hypotheses hold that very small antigenic particles penetrate the lung's defenses and come in contact with mast cells that interdigitate with the epithelium at the luminal surface of the central airways. The subsequent elaboration of mediators and cytokines then produces the sequence outlined above.

Pharmacologic Stimuli The drugs most commonly associated with the induction of acute episodes of asthma are aspirin, coloring agents such as tartrazine,b-adrenergic antagonists, and sulfiting agents. It is important to recognize drug-induced bronchial narrowing because its presence is often associated with great morbidity. Furthermore, death sometimes has followed the ingestion of aspirin (or other nonsteroidal anti-inflammatory agents) orb-adrenergic antagonists. The typical aspirin-sensitive respiratory syndrome primarily affects adults, although the condition may occur in childhood. This problem usually begins with perennial vasomotor rhinitis that is followed by a hyperplastic rhinosinusitis with nasal polyps. Progressive asthma then appears. On exposure to even very small quantities of aspirin, affected individuals typically develop ocular and nasal congestion and acute, often severe episodes of airways obstruction. The prevalence of aspirin sensitivity in patients with asthma varies from study to study, but many authorities feel that 10% is a reasonable figure. There is a great deal of cross reactivity between aspirin and other nonsteroidal anti-inflammatory compounds that inhibit prostaglandin G/H synthase 1 (cyclooxygenase type 1). Indomethacin, fenoprofen, naproxen, zomepirac sodium, ibuprofen, mefenamic acid, and phenylbutazone are particularly important in this regard. However, acetaminophen, sodium salicylate, choline salicylate, salicylamide, and propoxyphene are well tolerated. The exact frequency of cross reactivity to tartrazine and other dyes in aspirin-sensitive individuals with asthma is also controversial; again, 10% is the commonly accepted figure. This peculiar complication of aspirin-sensitive asthma is particularly insidious,

however, in that tartrazine and other potentially troublesome dyes are widely present in the environment and may be unknowingly ingested by sensitive patients.

Patients with aspirin sensitivity can be desensitized by daily administration of the drug. After this form of therapy, cross tolerance also develops to other nonsteroidal anti-inflammatory agents. The mechanism by which aspirin and other such drugs produce bronchospasm appears to be a chronic overexcretion of cysteinyl leukotrienes, which activate mast cells. The adverse reaction to aspirin can be inhibited with the use of leukotriene synthesis blockers or receptor antagonists.

Beta-adrenergic antagonists regularly obstruct the airways in individuals with asthma as well as in others with heightened airway reactivity and should be avoided by such individuals. Even the selective $beta_1$agents have this propensity, particularly at higher doses. In fact, the local use of $beta_1$blockers in the eye for the treatment of glaucoma has been associated with worsening asthma.

Sulfiting agents, such as potassium metabisulfite, potassium and sodium bisulfite, sodium sulfite, and sulfur dioxide, which are widely used in the food and pharmaceutical industries as sanitizing and preserving agents, also can produce acute airway obstruction in sensitive individuals. Exposure usually follows ingestion of food or beverages containing these compounds, e.g., salads, fresh fruit, potatoes, shellfish, and wine. Exacerbation of asthma has been reported after the use of sulfite-containing topical ophthalmic solutions, intravenous glucocorticoids, and some inhalational bronchodilator solutions. The incidence and mechanism of action of this phenomenon are unknown. When suspected, the diagnosis can be confirmed by either oral or inhalational provocations.

Environment and Air Pollution (See also Chap. 254) Environmental causes of asthma are usually related to climatic conditions that promote the concentration of atmospheric pollutants and antigens. These conditions tend to develop in heavily industrial or densely populated urban areas and are frequently associated with thermal inversions or other situations creating stagnant air masses. In these circumstances, although the general population can develop respiratory symptoms, patients with asthma and other respiratory diseases tend to be more severely affected. The air pollutants known to have this effect are ozone, nitrogen dioxide, and sulfur dioxide. Sulfur dioxide needs to be present in high concentrations and produces its greatest effects during periods of high ventilation. In some regions of North America, seasonal concentrations of airborne antigens such as pollen can rise high enough to result in epidemics of asthma admissions to hospitals and an increase in the death rate. These events may be ameliorated by treating patients prophylactically with anti-inflammatory drugs before the allergy season begins.

Occupational Factors (See alsoChap. 254) Occupation-related asthma is a significant health problem, and acute and chronic airway obstruction has been reported to follow exposure to a large number of compounds used in many types of industrial processes. Bronchoconstriction can result from working with or being exposed to *metal salts* (e.g., platinum, chrome, and nickel), *wood and vegetable dusts* (e.g., those of oak, western red cedar, grain, flour, castor bean, green coffee bean, mako, gum acacia, karay gum, and tragacanth), *pharmaceutical agents* (e.g., antibiotics, piperazine, and cimetidine),

industrial chemicals and plastics (e.g., toluene diisocyanate, phthalic acid anhydride, trimellitic anhydride, persulfates, ethylenediamine, *p*-phenylenediamine, and various dyes), *biologic enzymes* (e.g., laundry detergents and pancreatic enzymes), and *animal and insect dusts, serums, and secretions*. It is important to recognize that exposure to sensitizing chemicals, particularly those used in paints, solvents, and plastics, also can occur during leisure or non-work-related activities.

There seem to be three underlying mechanisms for this airway obstruction: (1) In some cases, the offending agent results in the formation of a specific IgE, and the cause seems immunologic (the immunologic reaction can be immediate, late, or dual); (2) in other cases, the substance causes a direct liberation of bronchoconstrictor substances; and (3) in other instances, the substance causes direct or reflex stimulation of the airways of individuals with either latent or frank asthma. If the occupational agent causes an immediate or dual immunologic reaction, the history is similar to that which occurs with exposure to other antigens. Often, however, patients will give a characteristic cyclic history. They are well when they arrive at work, and symptoms develop toward the end of the shift, progress after the work site is left, and then regress. Absence from work during weekends or vacations brings about remission. Frequently, there are similar symptoms in fellow employees.

Infections Respiratory infections are the most common of the stimuli that evoke acute exacerbations of asthma. Well-controlled investigations have demonstrated that respiratory viruses and not bacteria or allergy to microorganisms are the major etiologic factors. In young children, the most important infectious agents are respiratory syncytial virus and parainfluenza virus. In older children and adults, rhinovirus and influenza virus predominate as pathogens. Simple colonization of the tracheobronchial tree is insufficient to evoke acute episodes of bronchospasm, and attacks of asthma occur only when symptoms of an ongoing respiratory tract infection are, or have been, present. Viral infections can actively and chronically destabilize asthma, and they are perhaps the only stimuli that can produce constant symptoms for weeks. The mechanism by which viruses induce exacerbations of asthma may be related to the production of T cell-derived cytokines that potentiate the infiltration of inflammatory cells into already susceptible airways.

Exercise Exercise is a very common precipitant of acute episodes of asthma. This stimulus differs from other naturally occurring provocations, such as antigens, viral infections, and air pollutants, in that it does not evoke any long-term sequelae, nor does it increase airway reactivity. Exercise can be made to provoke bronchospasm in every patient with asthma, and in some it is the only trigger that produces symptoms. When such patients are followed for sufficient periods, however, they often develop recurring episodes of airway obstruction independent of exercise; thus, the onset of this problem frequently is the first manifestation of the full-blown asthmatic syndrome. The critical variables that determine the severity of the postexertional airway obstruction are the levels of ventilation achieved and the temperature and humidity of the inspired air. The higher the ventilation and the lower the heat content of the air, the greater the response. For the same inspired air conditions, running produces a more severe attack of asthma than walking. Conversely, for a given task, the inhalation of cold air markedly enhances the response, while warm, humid air blunts or abolishes it. Consequently, activities such as ice hockey, cross-country skiing, and ice skating are more provocative than is

swimming in an indoor, heated pool. The mechanism by which exercise produces obstruction may be related to a thermally produced hyperemia and engorgement of the microvasculature of the bronchial wall and does not appear to involve smooth-muscle contraction.

Emotional Stress Abundant objective data demonstrate that psychological factors can interact with the asthmatic diathesis to worsen or ameliorate the disease process. The pathways and nature of the interactions are complex but are operational to some extent in almost half the patients studied. Changes in airway caliber seem to be mediated through modification of vagal efferent activity, but endorphins also may play a role. The most frequently studied variable has been that of suggestion, and the weight of current evidence indicates that it can be quite important in selected individuals with asthma. When psychically responsive individuals are given the appropriate suggestion, they can actually decrease or increase the pharmacologic effects of adrenergic and cholinergic stimuli on their airways. The extent to which psychological factors participate in the induction and/or continuation of any given acute exacerbation is not established but probably varies from patient to patient and in the same patient from episode to episode.

PATHOLOGY

In a patient who has died of acute asthma, the most striking feature of the lungs at necropsy is their gross overdistention and failure to collapse when the pleural cavities are opened. When the lungs are cut, numerous gelatinous plugs of exudate are found in most of the bronchial branches down to the terminal bronchioles. Histologic examination shows hypertrophy of the bronchial smooth muscle, hyperplasia of mucosal and submucosal vessels, mucosal edema, denudation of the surface epithelium, pronounced thickening of the basement membrane, and eosinophilic infiltrates in the bronchial wall. There is an absence of any of the well-recognized forms of destructive emphysema.

PATHOPHYSIOLOGY

The pathophysiologic hallmark of asthma is a reduction in airway diameter brought about by contraction of smooth muscle, vascular congestion, edema of the bronchial wall, and thick, tenacious secretions. The net result is an increase in airway resistance, a decrease in forced expiratory volumes and flow rates, hyperinflation of the lungs and thorax, increased work of breathing, alterations in respiratory muscle function, changes in elastic recoil, abnormal distribution of both ventilation and pulmonary blood flow with mismatched ratios, and altered arterial blood gas concentrations. Thus, although asthma is considered to be primarily a disease of airways, virtually all aspects of pulmonary function are compromised during an acute attack. In addition, in very symptomatic patients there frequently is electrocardiographic evidence of right ventricular hypertrophy and pulmonary hypertension. When a patient presents for therapy, his or her forced vital capacity tends to be £50% of normal. The 1-s forced expiratory volume (FEV_1) averages 30% or less of predicted, while the maximum and minimum midexpiratory flow rates are reduced to 20% or less of expected. In keeping with the alterations in mechanics, the associated air trapping is substantial. In acutely ill patients, residual volume (RV) frequently approaches 400% of normal, while functional residual capacity doubles. The patient tends to report that the attack has ended clinically

when the RV has fallen to 200% of its predicted value and the FEV_1 reaches 50% of the predicted level.

Hypoxia is a universal finding during acute exacerbations, but frank ventilatory failure is relatively uncommon, being observed in 10 to 15% of patients presenting for therapy. Most individuals with asthma have hypocapnia and a respiratory alkalosis. In acutely ill patients, the finding of a normal arterial carbon dioxide tension tends to be associated with quite severe levels of obstruction. Consequently, when found in a symptomatic individual, it should be viewed as representing impending respiratory failure, and the patient should be treated accordingly. Equally, the presence of metabolic acidosis in the setting of acute asthma signifies severe obstruction. Usually, there are no clinical counterparts to the derangements in blood gases. Cyanosis is a very late sign. Hence, a dangerous level of hypoxia can go undetected. Likewise, signs attributable to carbon dioxide retention, such as sweating, tachycardia, and wide pulse pressure, or to acidosis, such as tachypnea, tend not to be of great value in predicting the presence of hypercapnia or hydrogen ion excess in individual patients, because they are too frequently seen in anxious patients with more moderate disease. Trying to judge the state of an acutely ill patient's ventilatory status on clinical grounds alone can be extremely hazardous, and clinical indicators should not be relied on with any confidence. Therefore, in patients with suspected alveolar hypoventilation, arterial blood gas tensions must be measured.

CLINICAL FEATURES

The symptoms of asthma consist of a triad of dyspnea, cough, and wheezing, the last often being regarded as the *sine qua non*. In its most typical form, asthma is an episodic disease, and all three symptoms coexist. At the onset of an attack, patients experience a sense of constriction in the chest, often with a nonproductive cough. Respiration becomes audibly harsh, wheezing in both phases of respiration becomes prominent, expiration becomes prolonged, and patients frequently have tachypnea, tachycardia, and mild systolic hypertension. The lungs rapidly become overinflated, and the anteroposterior diameter of the thorax increases. If the attack is severe or prolonged, there may be a loss of adventitial breath sounds, and wheezing becomes very high pitched. Furthermore, the accessory muscles become visibly active, and a paradoxical pulse often develops. These two signs are extremely valuable in indicating the severity of the obstruction. In the presence of either, pulmonary function tends to be significantly more impaired than in their absence. It is important to note that the development of a paradoxical pulse depends on the generation of large negative intrathoracic pressures. Thus, if the patient's breathing is shallow, this sign and/or the use of accessory muscles could be absent even though obstruction is quite severe. The other signs and symptoms of asthma only imperfectly reflect the physiologic alterations that are present. Indeed, if the disappearance of subjective complaints or even of wheezing is used as the end point at which therapy for an acute attack is terminated, an enormous reservoir of residual disease will be missed.

The end of an episode is frequently marked by a cough that produces thick, stringy mucus, which often takes the form of casts of the distal airways (Curschmann's spirals) and, when examined microscopically, often shows eosinophils and Charcot-Leyden crystals. In extreme situations, wheezing may lessen markedly or even disappear,

cough may become extremely ineffective, and the patient may begin a gasping type of respiratory pattern. These findings imply extensive mucus plugging and impending suffocation. Ventilatory assistance by mechanical means may be required. Atelectasis due to inspissated secretions occasionally occurs with asthmatic attacks. Spontaneous pneumothorax and/or pneumomediastinum occur but are rare.

Less typically, a patient with asthma may complain of intermittent episodes of nonproductive cough or exertional dyspnea. Unlike other individuals with asthma, when these patients are examined during symptomatic periods, they tend to have normal breath sounds but may wheeze after repeated forced exhalations and/or may show ventilatory impairments when tested in the laboratory. In the absence of both these signs, a bronchoprovocation test may be required to make the diagnosis.

DIFFERENTIAL DIAGNOSIS

The differentiation of asthma from other diseases associated with dyspnea and wheezing is usually not difficult, particularly if the patient is seen during an acute episode. The physical findings and symptoms listed above and the history of periodic attacks are quite characteristic. A personal or family history of allergic diseases such as eczema, rhinitis, or urticaria is valuable contributory evidence. An extremely common feature of asthma is nocturnal awakening with dyspnea and/or wheezing. In fact, this phenomenon is so prevalent that its absence raises doubt about the diagnosis. *Upper airway obstruction by tumor* or *laryngeal edema* can occasionally be confused with asthma. Typically, a patient with such a condition will present with stridor, and the harsh respiratory sounds can be localized to the area of the trachea. Diffuse wheezing throughout both lung fields is usually absent. However, differentiation can sometimes be difficult, and indirect laryngoscopy or bronchoscopy may be required. Asthma-like symptoms have been described in patients with glottic dysfunction. These individuals narrow their glottis during inspiration and expiration, producing episodic attacks of severe airway obstruction. Occasionally, carbon dioxide retention develops. However, unlike asthma, the arterial oxygen tension is well preserved, and the alveolar-arterial gradient for oxygen narrows during the episode, instead of widening as with lower airway obstruction. To establish the diagnosis of glottic dysfunction, the glottis should be examined when the patient is symptomatic. Normal findings at such a time exclude the diagnosis; normal findings during asymptomatic periods do not.

Persistent wheezing localized to one area of the chest in association with paroxysms of coughing indicates *endobronchial disease* such as foreign-body aspiration, a neoplasm, or bronchial stenosis.

The signs and symptoms of *acute left ventricular failure* occasionally mimic asthma, but the findings of moist basilar rales, gallop rhythms, blood-tinged sputum, and other signs of heart failure (Chap. 232) allow the appropriate diagnosis to be reached.

Recurrent episodes of bronchospasm can occur with *carcinoid tumors* (Chap. 93), *recurrent pulmonary emboli* (Chap. 261), and *chronic bronchitis* (Chap. 258). In chronic bronchitis there are no true symptom-free periods, and one can usually obtain a history of chronic cough and sputum production as a background on which acute attacks of wheezing are superimposed. Recurrent emboli can be very difficult to separate from

asthma. Frequently, patients with this condition present with episodes of breathlessness, particularly on exertion, and they sometimes wheeze. Pulmonary function studies may show evidence of peripheral airway obstruction (Chap. 250); when these changes are present, lung scans also may be abnormal. The therapeutic response to bronchodilators and to the institution of anticoagulant therapy may be helpful, but pulmonary angiography may be necessary to establish the correct diagnosis.

Eosinophilic pneumonias (Chap. 253) are often associated with asthmatic symptoms, as are various chemical pneumonias and exposures to insecticides and cholinergic drugs. Bronchospasm occasionally is a manifestation of *systemic vasculitis* with pulmonary involvement.

DIAGNOSIS

The diagnosis of asthma is established by demonstrating reversible airway obstruction. *Reversibility* is traditionally defined as a 15% or greater increase in FEV_1after two puffs of a b-adrenergic agonist. When the spirometry results are normal at presentation, the diagnosis can be made by showing heightened airway responsiveness to challenges with histamine, methacholine, or isocapnic hyperventilation of cold air. Once the diagnosis is confirmed, the course of the illness and the effectiveness of therapy can be followed by measuring peak expiratory flow rates (PEFRs) at home and/or the FEV_1 in the laboratory. Positive wheal-and-flare reactions to skin tests can be demonstrated to various allergens, but such findings do not necessarily correlate with the intrapulmonary events. Sputum and blood eosinophilia and measurement of serum IgE levels are also helpful but are not specific for asthma. Chest roentgenograms showing hyperinflation are also nondiagnostic.

TREATMENT

Elimination of the causative agent(s) from the environment of an allergic individual with asthma is the most successful means available for treating this condition (for details on avoidance, see Chap. 310). Desensitization or immunotherapy with extracts of the suspected allergens has enjoyed widespread favor, but controlled studies are limited and have not proved it to be highly effective.

Drug Treatment The available agents for treating asthma can be divided into two general categories: drugs that inhibit smooth muscle contraction, i.e., the so-called "quick relief medications" (beta-adrenergic agonists, methylxanthines, and anticholinergics) and agents that prevent and/or reverse inflammation, i.e., the "long-term control medications" (glucocorticoids, leukotriene inhibitors and receptor antagonists, and mast cell-stabilizing agents).

Adrenergic Stimulants The drugs in this category consist of the catecholamines, resorcinols, and saligenins. These agents are analogues and produce airway dilation through stimulation of beta-adrenergic receptors and activation of G proteins with the resultant formation of cyclic adenosine monophosphate (AMP). They also decrease release of mediators and improve mucociliary transport. The catecholamines available for clinical use are epinephrine, isoproterenol, and isoetharine. As a group, these

compounds are short-acting (30 to 90 min) and are effective only when administered by inhalational or parenteral routes. Epinephrine and isoproterenol are notb_2-selective and have considerable chronotropic and inotropic cardiac effects. Epinephrine also has substantial alpha-stimulating effects. The usual dose is 0.3 to 0.5 mL of a 1:1000 solution administered subcutaneously. Isoproterenol is devoid of alpha activity and is the most potent agent of this group. It is usually administered in a 1:200 solution by inhalation. Isoetharine is the mostb_2-selective compound of this class, but it is a relatively weak bronchodilator. It is employed as an aerosol and supplied as a 1% solution. The use of these agents in treating asthma has been superceded by longer acting selective$b_2$agonists.

The commonly used resorcinols are metaproterenol, terbutaline, and fenoterol, and the most widely known saligenin is albuterol (salbutamol). With the exception of metaproterenol, these drugs are highly selective for the respiratory tract and virtually devoid of significant cardiac effects except at high doses. Their major side effect is tremor. They are active by all routes of administration, and because their chemical structures allow them to bypass the metabolic processes used to degrade the catecholamines, their effects are relatively long-lasting (4 to 6 h). Differences in potency and duration among agents can be eliminated by adjusting doses and/or administration schedules.

Inhalation is the preferred route of administration because it allows maximal bronchodilation with fewer side effects. In the past it was fashionable to treat episodes of severe asthma with intravenous sympathomimetics such as isoproterenol. This approach no longer appears justifiable. Isoproterenol infusions clearly can induce myocardial damage, and even for theb_2-selective agents such as terbutaline and albuterol, intravenous administration offers no advantages over the inhaled route.

Salmeterol is a very long-lasting (9 to 12 h) congener of albuterol. When given every 12 h, it is effective in providing sustained symptomatic relief. It is particularly helpful for conditions such as nocturnal and exercise-induced asthma. It is not recommended for the treatment of acute episodes because of its relatively slow onset of action (approximately 30 min), nor is it intended as a rescue drug for breakthrough symptoms. In addition, its long half-life means that administration of extra doses can cause cumulative side effects.

Methylxanthines Theophylline and its various salts are medium-potency bronchodilators that work by increasing cyclicAMP by the inhibition of phosphodiesterase. The therapeutic plasma concentrations of theophylline traditionally have been thought to lie between 10 and 20 ug/mL. Some sources, however, recommend a lower target range between 5 and 15 ug/mL to avoid toxicity. The dose required to achieve the desired level varies widely from patient to patient owing to differences in the metabolism of the drug. Theophylline clearance, and thus the dosage requirement, is decreased substantially in neonates and the elderly and those with acute and chronic hepatic dysfunction, cardiac decompensation, and cor pulmonale. Clearance is also decreased during febrile illnesses. Clearance is increased in children. In addition, a number of important drug interactions can alter theophylline metabolism. Clearance falls with the concurrent use of erythromycin and other macrolide antibiotics, the quinolone antibiotics, and troleandomycin, allopurinol, cimetidine, and propranolol. It

rises with use of cigarettes, marijuana, phenobarbital, phenytoin, or any other drug that is capable of inducing hepatic microsomal enzymes.

For maintenance therapy, long-acting theophylline compounds are available and are usually given once or twice daily. The dose is adjusted on the basis of the clinical response with the aid of serum theophylline measurements. Single-dose administration in the evening reduces nocturnal symptoms and helps keep the patient complaint-free during the day. Aminophylline and theophylline are available for intravenous use. The recommendations for intravenous therapy in children aged 9 to 16 and in young adult smokers not currently receiving theophylline products are a loading dose of 6 mg/kg followed by an infusion of 1 mg/kg per hour for the next 12 h and then 0.8 mg/kg per hour thereafter. In nonsmoking adults, older patients, and those with cor pulmonale, congestive heart failure, and liver disease, the loading dose remains the same, but the maintenance dose is reduced to between 0.1 and 0.5 mg/kg per hour. In patients already receiving theophylline, the loading dose is frequently withheld or, in extreme situations, reduced to 0.5 mg/kg.

The most common side effects of theophylline are nervousness, nausea, vomiting, anorexia, and headache. At plasma levels greater than 30 ug/mL there is a risk of seizures and cardiac arrhythmias.

Anticholinergics Anticholinergic drugs such as atropine sulfate produce bronchodilation in patients with asthma, but their use is limited by systemic side effects. Nonabsorbable quaternary ammonium congeners (atropine methylnitrate and ipratropium bromide) have been found to be both effective and free of untoward effects. They may be of particular benefit for patients with coexistent heart disease, in whom the use of methylxanthines andb-adrenergic stimulants may be dangerous. The major disadvantages of the anticholinergics are that they are slow to act (60 to 90 min may be required before peak bronchodilation is achieved) and they are only of modest potency.

Glucocorticoids Glucocorticoids are the most potent and most effective anti-inflammatory medications available. Systemic or oral steroids are most beneficial in acute illness when severe airway obstruction is not resolving or is worsening despite intense optimal bronchodilator therapy, and in chronic disease when there has been failure of a previously optimal regimen with frequent recurrences of symptoms of increasing severity. Inhaled glucocorticoids are used in the long-term control of asthma.

Glucocorticoids are not bronchodilators and the correct dose to use in acute situations is a matter of debate. The available data indicate that very high doses do not offer any advantage over more conventional amounts. In the United States, a usual starting dose is 40 to 60 mg of methylprednisolone intravenously every 6 h. Since intravenous and oral administration produce the same effects, prednisone, 60 mg every 6 h, can be substituted. Clinical impressions suggest that smaller quantities may work as effectively, but there are no confirmatory data. In the United Kingdom and elsewhere, acute asthma both in and out of hospital is frequently treated with doses of prednisolone ranging from 30 to 40 mg given once daily. It should be emphasized that the effects of steroids in acute asthma are not immediate and may not be seen for 6 h or more after the initial administration. Consequently, it is mandatory to continue vigorous bronchodilator therapy during this interval. Irrespective of the regimen chosen, it is important to

appreciate that rapid tapering of glucocorticoids frequently results in recurrent obstruction. Most authorities recommend reducing the dose by one-half every third to fifth day after an acute episode. In situations in which it appears that continued steroid therapy will be needed, an alternate-day schedule should be instituted to minimize side effects. This is particularly important in children, since continuous glucocorticoid administration interrupts growth. Long-acting preparations such as dexamethasone should not be used in this approach, for they defeat the purpose of alternate-day schedules by causing prolonged suppression of the pituitary-adrenal axis. The availability of inhaled agents has all but eliminated the need for this form of therapy.

INHALED GLUCOCORTICOIDS These drugs are indicated in patients with persistent symptoms. The agents currently available in the United States are beclomethasone, budesonide, flunisolide, fluticasone propionate, and triamcinolone acetonide. Each has relative advantages and disadvantages, and they are not absolutely interchangeable on either a microgram or a per puff basis. However, all of these drugs share the ability to control inflammation, facilitate the long-term prevention of symptoms, and reduce the need for oral glucocorticoids.

There is no fixed dose of inhaled steroid that works for all patients. Requirements are dictated by the response of the individual and wax and wane in concert with progression of the disease. Generally, the worse the patient's condition, the more inhaled steroid is needed to gain control. Once achieved, however, remission can often be maintained with quantities as low as one or two puffs/day. Inhaled steroids can take up to a week or more to produce improvements; consequently, in rapidly deteriorating situations, it is best to prescribe oral preparations and initiate inhaled drugs as the dose of the former is reduced. In less emergent circumstances, the quantity of inhaled drug can be increased up to 2 to 2.5 times the recommended starting doses. It is critical to remember that the side effects increase in proportion to the dose-time product. In addition to thrush and dysphonia, the increased systemic absorption that accompanies larger doses of inhaled steroids has been reported to produce adrenal suppression, cataract formation, decreased growth in children, interference with bone metabolism, and purpura. As is the case with oral agents, suppression of inflammation, per se, cannot be relied upon to provide optimal results. It is essential to continue adrenergic or methylxanthine bronchodilators if the patient's disease is unstable.

Mast Cell-Stabilizing Agents Cromolyn sodium and nedocromil sodium do not influence airway tone. Their major therapeutic effect is to inhibit the degranulation of mast cells, thereby preventing the release of the chemical mediators of anaphylaxis.

Cromolyn sodium and nedocromil, like the inhaled steroids, improve lung function, reduce symptoms, and lower airway reactivity in persons with asthma. They are most effective in atopic patients who have either seasonal disease or perennial airway stimulation. A therapeutic trial of two puffs four times daily for 4 to 6 weeks frequently is necessary before the beneficial effects of the drug appear. Unlike steroids, nedocromil and cromolyn sodium, when given prophylactically, block the acute obstructive effects of exposure to antigen, industrial chemicals, exercise, or cold air. With antigen, the late response is also abolished. Therefore, a patient who has intermittent exposure to either antigenic or nonantigenic stimuli that provoke acute episodes of asthma need not use these drugs continuously but instead can obtain protection by taking the drug only 15 to

20 min before contact with the precipitant.

Leukotriene Modifiers As mentioned earlier, the cysteinyl leukotrienes (LTC_4, LTD_4, and LTE_4) produce many of the critical elements of asthma, and drugs have been developed to either reduce the synthesis of all of the leukotrienes by inhibiting 5-lipoxygenase (5-LO), the enzyme involved in their production, or competitively antagonizing the principal moiety (LTD_4). Zileuton is the only 5-lipoxygenase synthesis inhibitor that is available in the United States. It is a modest bronchodilator that reduces asthma morbidity, provides protection against exercise-induced asthma, and diminishes nocturnal symptoms, but it has limited effectiveness against allergens. Hepatic enzyme levels can be elevated after its use, and there are significant interactions with other drugs metabolized in the liver. The LTD_4receptor antagonists (zafirlukast and montelukast) have therapeutic and toxicologic profiles similar to that of zileuton but are long acting and permit twice to single daily dose schedules.

This class of drugs does not appear to be uniformly effective in all patients with asthma. Although precise figures are lacking, most authorities put the number of positive responders at less than 50%. As yet, there is no way of determining prospectively who will benefit, so clinical trials are required. Typically, if there is no improvement after one month, treatment can be discontinued.

Miscellaneous Agents It has been suggested that steroid-dependent patients might benefit from the use of immunosuppressant agents such as methotrexate or gold salts. The effects of these agents on steroid dosage and disease activity are minor, and side effects can be considerable. Consequently, this form of treatment can be viewed only as experimental. Opiates, sedatives, and tranquilizers should be absolutely avoided in the acutely ill patient with asthma because the risk of depressing alveolar ventilation is great, and respiratory arrest has been reported to occur shortly after their use. Admittedly, most individuals are anxious and frightened, but experience has shown that they can be calmed equally well by the physician's presence and reassurances. b-Adrenergic blockers and parasympathetic agonists are contraindicated because they can cause marked deterioration in lung function.

Expectorants and mucolytic agents have enjoyed great vogue in the past, but they do not add significantly to the treatment of the acute or chronic phases of this disease. Mucolytic agents such as acetylcysteine may actually produce bronchospasm when administered to susceptible patients with asthma. This effect can be overcome by aerosolizing them in solution with a b-adrenergic agent. The use of intravenous fluids in the treatment of acute asthma also has been advocated. There is little evidence that this adjunct hastens recovery. Nonstandard bronchodilators, such as intravenous magnesium sulfate, for the treatment of acute asthma attacks are not yet warranted in clinical practice because of the controversy surrounding their efficacy.

Special Instructions The treatment of patients with asthma who have coexisting conditions such as heart disease or pregnancy does not differ materially from that outlined above. Therapy with inhaledb_2-selective and anti-inflammatory agents is the mainstay. The lowest doses of adrenergics that produce the desired effects should be used.

FRAMEWORK FOR MANAGEMENT

Emergency Situations The most effective treatment for acute episodes of asthma requires a systematic approach based on the aggressive use of sympathomimetic agents and serial monitoring of key indices of improvement. Reliance on empirism and subjective assessment is no longer acceptable. Multiple inhalations of a short-acting sympathomimetic, such as albuterol, are the cornerstone of most regimens. These drugs provide three to four times more relief than does intravenous aminophylline. Anticholinergic drugs are not first-line therapy because of their long lag time to onset (~ 30 to 40 min) and their relatively modest bronchodilator properties. In emergency situations,$b_2$agonists can be given every 20 min by handheld nebulizer for 2 to 3 doses. The optimum cumulative dose of albuterol appears to lie between 5 and 10 mg. It does not matter how the adrenergic agonists are inhaled. Treatment with albuterol administered by jet nebulizer, metered dose inhaler, or dry powder inhaler all provide equal resolution in acute situations. Aminophylline or ipratropium can be added to the regimen after the first hour in an attempt to speed resolution. Recent studies in a large series of patients demonstrate that$b_2$agonists alone terminate attacks in approximately two-thirds of patients, and that another 5 to 10% benefit from a methylxanthine or ipratropium in combination with a sympathomimetic. The remainder have a poor acute response to all forms of therapy.

Acute episodes of bronchial asthma are one of the most common respiratory emergencies seen in the practice of medicine, and it is essential that the physician recognize which episodes of airway obstruction are life-threatening and which patients demand what level of care. These distinctions can be made readily by assessing selected clinical parameters in combination with measures of expiratory flow and gas exchange. The presence of a paradoxical pulse, use of accessory muscles, and marked hyperinflation of the thorax signify severe airways obstruction, and failure of these signs to remit promptly after aggressive therapy mandates objective monitoring of the patient with measurements of arterial blood gases and the peak expiratory flow rate (PEFR) or FEV_1.

In general, there is a correlation between the severity of the obstruction with which the patient presents and the time it takes to resolve it. Those individuals with the most impairment typically require the most extensive therapy for resolution. If thePEFR or FEV_1 is equal to or less than 20% of predicted on presentation and does not double within an hour of receiving the preceding therapy, the patient is likely to require extensive treatment including glucocorticoids before the obstruction dissipates. This group represents approximately 20% of all the patients who present for acute care. They generally require 3 to 4 days of inpatient treatment before becoming asymptomatic. In such individuals, if the clinical signs of a paradoxical pulse and accessory muscle use are diminishing, and/or if PEFR is increasing, there is no need to change medications or doses; the patient need only be followed. However, if the PEFR falls by more than 20% of its previous value or if the magnitude of the pulsus paradoxicus is increasing, serial measures of arterial blood gases are required, as well as a reconsideration of the therapeutic modalities being employed. If the patient has hypocarbia, one can afford to continue the current approaches a while longer. On the other hand, if the Pa_{CO_2} is within the normal range or is elevated, the patient should be monitored in an intensive care setting, and therapy should be intensified to reverse or

arrest the patient's respiratory failure.

Chronic Treatment The goal of chronic therapy is to achieve a stable, asymptomatic state with the best pulmonary function possible using the least amount of medication. The first step is to educate patients to function as partners in their management. The severity of the illness needs to be assessed and monitored with objective measures of lung function. Asthma triggers should be avoided or controlled, and plans should be made for both chronic management and treatment of exacerbations. Regular follow-up care is mandatory. With respect to pharmacologic interventions, in general, the simplest approach works best. Infrequent symptoms require only the use of an inhaled sympathomimetic on an "as needed" basis. When the disease worsens, as manifested by nocturnal awakenings and daytime symptoms, inhaled steroids and/or mast cell-stabilizing agents should be added. If symptoms do not abate, the dose of inhaled steroids can be increased. An upper limit has not yet been established, but side effects of glucocorticoid excess begin to appear more frequently when the dose exceeds 2.0 mg/d. Persistent asthma complaints can be treated with long-acting inhaled$b_2$agonists, sustained-release theophylline, and/or parasympatholytics. In patients with recurrent or perennial symptoms and unstable lung function, oral steroids in a single daily dose are added to the regimen. Once control is reached and sustained for several weeks, a step-down reduction in therapy should be undertaken, beginning with the most toxic drug, to find the minimum amount of medication required to keep the patient well. During this process, thePEFRshould be monitored and medication adjustments should be based on objective changes in lung function as well as on the patient's symptoms.

PROGNOSIS AND CLINICAL COURSE

The mortality rate from asthma is small. The most recent figures indicate fewer than 5000 deaths per year out of a population of approximately 10 million patients at risk. Death rates, however, appear to be rising in inner-city areas where there is limited availability of health care.

Information on the clinical course of asthma suggests a good prognosis particularly for those whose disease is mild and develops in childhood. The number of children who still have asthma 7 to 10 years after the initial diagnosis varies from 26 to 78%, averaging 46%; however, the percentage who continue to have severe disease is relatively low (6 to 19%).

Although there are reports of patients with asthma developing irreversible changes in lung function, these individuals frequently have comorbid stimuli such as cigarette smoking that could account for these findings. Even when untreated, individuals with asthma do not continuously move from mild to severe disease with time. Rather, their clinical course is characterized by exacerbations and remissions. Some studies suggest that spontaneous remissions occur in approximately 20% of those who develop the disease as adults and that 40% or so can be expected to experience improvement, with less frequent and severe attacks, as they grow older.

(Bibliography omitted in Palm version)

Back to Table of Contents

253. HYPERSENSITIVITY PNEUMONITIS AND PULMONARY INFILTRATES WITH EOSINOPHILIA - *Joel N. Kline*, *Gary W. Hunninghake*

HYPERSENSITIVITY PNEUMONITIS

Hypersensitivity pneumonitis (HP), or extrinsic allergic alveolitis, is an immunologically induced inflammatory disorder of the lung parenchyma, involving alveolar walls and terminal airways, secondary to repeated inhalation of a variety of organic agents by a susceptible host. Causes of HP are typically designated with colorful names denoting the occupational or avocational risk associated with the disease; "farmer's lung" is the term most commonly used for HP due to inhalation of antigens present in moldy hay, such as thermophilic actinomyces, *Micropolyspora faeni*, and *Aspergillus* species. The prevalence of HP is unknown but varies with the environmental exposure and the specific antigen involved. The prevalence of farmer's lung among Wisconsin dairy farmers has been reported as 4.2 per 1000. The diagnosis of HP requires a constellation of clinical, radiographic, physiologic, pathologic, and immunologic criteria, each of which is rarely pathognomonic alone; and the preferred treatment is avoidance of the causative antigen when practical.

ETIOLOGY

Agents implicated as causes ofHPinclude those listed in Table 253-1. Many cases of HP occurring in various occupations involve exposure to similar agents, particularly the thermophilic actinomycetes. In the United States, the most common types of HP are farmer's lung, bird fancier's lung, and chemical worker's lung. In *farmer's lung*, inhalation of proteins such as thermophilic bacteria and fungal spores that are present in moldy bedding and feed are most commonly responsible for the development of HP. These antigens are probably also responsible for the etiology of *mushroom worker's disease* (moldy composted growth medium is the source of the proteins) and *bagassosis* (moldy sugar cane is the source). *Bird fancier's lung* (and the related disorders of duck fever, turkey handler's lung, and dove pillow's lung) is a response to inhalation of bird proteins from feathers and droppings. *Chemical worker's lung* is an example of how simple chemicals, such as isocyanates, may also cause immune-mediated diseases. In this case, antihapten antibodies may be responsible for the development of HP.

PATHOGENESIS

The finding that precipitating antibodies against extracts of moldy hay were demonstrable in most patients with farmer's lung led to the early conclusion that HPwas an immune-complex-mediated reaction. Subsequent investigations of HP in humans and animal models provided evidence for the importance of cell-mediated hypersensitivity. The very early (acute) reaction is characterized by an increase in polymorphonuclear leukocytes in the alveoli and small airways. This early lesion is followed by an influx of mononuclear cells into the lung and the formation of granulomas that appear to be the result of a classic delayed (T cell mediated) hypersensitivity reaction to repeated inhalation of antigen and adjuvant-active materials. Recent studies in animal models suggest that the disease is mediated as a classic T_H1 cell-mediated immune response to antigen.

Bronchoalveolar lavage in patients with HPconsistently demonstrates an increase in the number of T lymphocytes in lavage fluid (a finding that is also observed in patients with other granulomatous lung disorders). Patients with recent or continual exposure to antigen may have an increase in the number of polymorphonuclear leukocytes in lavage fluid. Increased numbers of mast cells have also been reported. In most patients examined during recovery from acute disease, the T lymphocytes in lavage fluid are predominantly the CD8+ T cell subset. In patients with very recent exposure to antigen, however, the numbers of CD4+ T cells may increase in lavage fluid. Similar findings may be present in similarly exposed, asymptomatic individuals. These observations and others in animal models suggest that there is an active modulation of granuloma formation in the lung by immunoregulatory T cells and associated cytokines in this disorder.

CLINICAL PRESENTATION

The clinical picture is that of an interstitial pneumonitis, although it varies from patient to patient and seems related to the frequency and intensity of exposure to the causative antigen and perhaps other host factors. The presentation can be *acute, subacute*, or *chronic*. In the *acute form*, symptoms such as cough, fever, chills, malaise, and dyspnea may occur 6 to 8 h after exposure to the antigen and usually clear within a few days if there is no further exposure to antigen. The *subacute form* often appears insidiously over a period of weeks marked by cough and dyspnea and may progress to cyanosis and severe dyspnea requiring hospitalization. In some patients, a subacute form of the disease may persist after an acute presentation of the disorder, especially if there is continued exposure to antigen. In most patients with the acute or subacute form ofHP, the symptoms, signs, and other manifestations of HP disappear within days, weeks, or months if the causative agent is no longer inhaled. Transformation to a chronic form of the disease may occur in patients with continued antigen exposure, but the frequency of such progression is uncertain. The *chronic form* of HP may be clinically indistinguishable from pulmonary fibrosis due to a wide variety of causes. Physical examination may reveal clubbing. This stage may progressively worsen, resulting in dependence on supplemental O_2, pulmonary hypertension, and death from respiratory failure. An indolent gradually progressive form of the disease can be associated with cough and exertional dyspnea without a prior history consistent with acute or subacute manifestations. Such a gradual onset frequently occurs with low-dose exposure to the antigen.

Because strict definitions of acute, subacute, and chronic stages ofHP have not been generally agreed on, interpretation of epidemiologic and clinical studies can be difficult. Therefore, it has been proposed that HP be described as recently diagnosed, recurrent or progressive, or residual disease. For these categories, required diagnostic criteria include the presence of an appropriate exposure; exertional dyspnea; inspiratory crackles; and, if performed, lymphocytic alveolitis on bronchoalveolar lavage. Supportive criteria include recurrent febrile episodes, radiographic infiltrates, diminished pulmonary diffusing capacity, precipitating antibodies to appropriate antigens, histopathologic demonstration of granulomas, and improvement in symptoms with avoidance of exposure.

DIAGNOSIS

After acute exposure to antigen, neutrophilia and lymphopenia are frequently present. Eosinophilia is not a feature. All forms of the disease may be associated with elevations in erythrocyte sedimentation rate, C-reactive protein, rheumatoid factor, and serum immunoglobulins. Antinuclear antibodies are rarely present and appear to have no pathogenic role. Examination for *serum precipitins* against suspected antigens, such as those listed inTable 253-1, is an important part of the diagnostic workup and should be performed on any patient with interstitial lung disease, especially if a suggestive exposure history is elicited. If found, precipitins indicate sufficient exposure to the causative agent for generation of an immunologic response. The diagnosis ofHP is not established solely by the presence of precipitins, however, as precipitins are found in sera of many individuals exposed to appropriate antigens who demonstrate no other evidence of HP. False-negative results may occur because of poor-quality antigens or an inappropriate choice of antigens. Extraction of antigens from the suspected source may at times be helpful.

No specific or distinctive *chest roentgenogram* occurs inHP. It can be normal even in symptomatic patients. The acute or subacute phase may be associated with poorly defined, patchy, or diffuse infiltrates or with discrete, nodular infiltrates (Fig. 253-1). In the chronic phase, the chest x-ray usually shows a diffuse reticulonodular infiltrate. Honeycombing may eventually develop as the condition progresses. Apical sparing is common, suggesting that disease severity correlates with inhaled antigen load, but no particular distribution or pattern is classic for HP. Abnormalities rarely seen in HP include pleural effusion or thickening, and hilar adenopathy. High-resolution chest computed tomography (CT) has been reported to show a characteristic constellation of abnormalities, including (1) global lung involvement with increased lung density, (2) prominence of medium-sized bronchial walls, (3) patchy air space opacification with reticular and nodular patterns and midzone prominence, and (4) absence of hilar lymph node enlargement. No pathognomonic CT features of HP have been described (Fig. 253-2).

Pulmonary function studies in all forms ofHP may show a restrictive or obstructive pattern with loss of lung volumes, impaired diffusion capacity, decreased compliance, and an exercise-induced hypoxemia. Resting hypoxemia may also be found. Bronchospasm and bronchial hyperreactivity are sometimes found in acute HP. With antigen avoidance, the pulmonary function abnormalities are usually reversible, but they may gradually increase in severity or may occur rapidly after acute or subacute exposure to antigen.

Bronchoalveolar lavage is used in some centers to aid in diagnostic evaluation. A marked lymphocytic alveolitis on bronchoalveolar lavage is almost universal, although not pathognomonic. Lymphocytes typically have a decreased helper/suppressor ratio and are activated. Alveolar neutrophilia is also prominent acutely but tends to fade in the absence of recurrent exposure. Bronchoalveolar mastocytosis may correlate with disease activity. *Lung biopsy*, obtained through flexible bronchoscopy, open-lung procedures, or thoracoscopy, may be diagnostic. Although the histopathology is distinctive, it may not be pathognomonic ofHP. When the biopsy is taken during the active phase of disease, typical findings include an interstitial alveolar infiltrate consisting of plasma cells, lymphocytes, and occasional eosinophils and neutrophils,

usually with accompanying granulomas. Interstitial fibrosis may be present but most often is mild in earlier stages of the disease. Some degree of bronchiolitis is found in about half the cases, whereas vasculitis is not a feature of the disorder. The triad of mononuclear bronchiolitis; interstitial infiltrates of lymphocytes and plasma cells; and single, nonnecrotizing, and randomly scattered parenchymal granulomas without mural vascular involvement is consistent with but not specific for HP.

Inhalation challenge studies have been described as useful to differentiate betweenHP and other interstitial lung diseases. These tests should be performed in a center that specializes in provocation testing for reasons of both safety and accuracy. Moveover, because the antigens used for provocation testing are not standardized, interpretation of these tests is difficult. In general, these tests may be used to support a diagnosis of HP, but they are not sufficiently accepted to either confirm or deny the diagnosis. The lack of standardized, nonirritating antigens and of proven controlled protocols makes *skin testing* useful only for research purposes. Similarly, in vitro tests of cell-mediated (delayed) hypersensitivity have not consistently been shown to correlate with clinical HP and have no place in the routine diagnostic workup.

In summary, the diagnosis in most cases is established by (1) consistent history, physical findings, pulmonary function tests, and chest x-ray; (2) exposure to a recognized antigen; and (3) finding an antibody to that antigen. In a few circumstances, bronchoalveolar lavage and/or lung biopsy may be needed. Provocation tests may be useful but are not essential for the diagnosis.

DIFFERENTIAL DIAGNOSIS

ChronicHP may often be difficult to distinguish from a number of other interstitial lung disorders such as idiopathic pulmonary fibrosis, sarcoidosis, interstitial lung disease associated with a collagen vascular disorder, and drug-induced lung diseases. A negative history for use of relevant drugs and no evidence of a systemic disorder usually exclude the presence of drug-induced lung disease or a collagen vascular disorder. Bronchoalveolar lavage often shows predominance of neutrophils in idiopathic pulmonary fibrosis and a predominance of CD4+ lymphocytes in sarcoidosis. Hilar/paratracheal lymphadenopathy or evidence of multisystem involvement also favors the diagnosis of sarcoidosis. In some patients, a lung biopsy may be required to differentiate chronic HP from other interstitial diseases. The lung disease associated with acute or subacute HP may clinically resemble other disorders that present with systemic symptoms and recurrent pulmonary infiltrates, including the allergic bronchopulmonary mycoses and other eosinophilic pneumonias.

Eosinophilic pneumonia is often associated with asthma and is typified by peripheral eosinophilia; neither of these is a feature ofHP. Allergic bronchopulmonary aspergillosis (ABPA) is the most common example of the allergic bronchopulmonary mycoses and is sometimes confused with HP because of the presence of precipitating antibodies to *A. fumigatus*. ABPA is associated with allergic (atopic) asthma. Acute HP may be confused with *organic dust toxic syndrome* (ODTS), a condition that is more common than HP. ODTS follows heavy exposure to organic dusts and is characterized by transient fever and muscle aches, with or without dyspnea and cough. Serum precipitins are absent, and the chest x-ray is usually normal. Studies have shown no immunologic basis for

ODTS, and endotoxin is suspected to be involved in its pathogenesis. This distinction is important, as ODTS is a self-limited disorder without significant long-term sequellae, whereas continued antigen exposure in HP can result in permanent disability. Massive exposure to moldy silage may result in a syndrome termed *pulmonary mycotoxicosis*, or *atypical farmer's lung*, with fever, chills, and cough and the presence of pulmonary infiltrates within a few hours of exposure. No previous sensitization is required, and precipitins are absent to *Aspergillus*, the suspected causative agent.

TREATMENT

Because effective treatment depends largely on avoiding the antigen, identification of the causative agent and its source is essential. This identification is usually possible if the physician takes a careful environmental and occupational history or, if necessary, visits the patient's environment. The simplest way to avoid the incriminated agent is to remove the patient from the environment or the source of the agent from the patient's environment. This recommendation cannot be taken lightly when it completely changes the life-style or livelihood of the patient. In many cases, however, the source of exposure (birds, humidifiers) can easily be removed. If occupational exposure is involved, an initial attempt can be made at antigen avoidance maneuvers that are least disruptive to the patient's livelihood, which usually means avoiding areas associated with heavy exposure and wearing an appropriate mask. This will not protect against small-molecular-weight agents such as isocyanates, which require more elaborate respiratory systems. Pollen masks, personal dust respirators, airstream helmets, and ventilated helmets with a supply of fresh air are increasingly efficient means of purifying inhaled air. If symptoms recur or physiologic abnormalities progress in spite of these measures, then more effective measures to avoid antigen exposure must be pursued. Compromises with environmental control pertain primarily to the acute, recurrent, transient clinical form ofHP and must be accompanied by careful follow-up. Subacute forms are ordinarily the result of a heavy, sustained exposure. The chronic form typically results from low-grade or recurrent exposure over many months or years, and the lung disease may already be partially irreversible. These patients are usually advised to avoid completely all possible contact with the offending agent, although follow-up studies of individuals with farmer's lung and bird fancier's lung have found resolution of the disease despite continued exposure in some patients.

Patients with the *acute*, recurrent form ofHPusually recover without need for glucocorticoids. *Subacute* HP may be associated with severe symptoms and marked physiologic impairment and may continue to progress for several days despite hospitalization. Urgent establishment of the diagnosis and prompt institution of glucocorticoid treatment are indicated in such patients. Such therapy may also hasten recovery in patients with lesser involvement. Prednisone at a dosage of 1 mg/kg per day or its equivalent is continued for 7 to 14 days and then tapered over the ensuing 2 to 6 weeks at a rate that depends on the patient's clinical status. Patients with *chronic* HP may gradually recover without therapy after the institution of environmental control. In many patients, however, a trial of prednisone may be useful to obtain maximal reversibility of the lung disease. After initial prednisone therapy (1 mg/kg per day for 2 to 4 weeks), the drug is tapered to the lowest dosage that will maintain the functional status of the patient. Many patients will not require or benefit from long-term therapy if there is no further exposure to antigen. Available studies report no effect of

glucocorticoid therapy on long-term prognosis of farmer's lung.

PULMONARY INFILTRATES WITH EOSINOPHILIA

Pulmonary infiltrates with eosinophilia (PIE, eosinophilic pneumonias) include distinct individual syndromes characterized by eosinophilic pulmonary infiltrates and, commonly, peripheral blood eosinophilia. Since Loeffler's initial description of a transient, benign syndrome of migratory pulmonary infiltrates and peripheral blood eosinophilia of unknown cause, this group of disorders has been enlarged to include several diseases of both known and unknown etiology (Table 253-2). These diseases may be considered as putative hypersensitivity lung diseases but are not to be confused withHP(extrinsic allergic alveolitis), in which eosinophilia is not a feature. When an eosinophilic pneumonia is associated with bronchial asthma, it is important to determine if the patient has atopic asthma and has wheal-and-flare skin reactivity to *Aspergillus* or other relevant fungal antigens. If so, other criteria should be sought for diagnosis ofABPA(Table 253-3) or other, rarer examples of allergic bronchopulmonary mycosis such as those caused by *Penicillium*, *Candida*, *Curvularia*, or *Helminthosporium* spp. *A. fumigatus* is the most common cause of ABPA, although other *Aspergillus* species have also been implicated. ABPA has been reported to complicate cystic fibrosis. The chest roentgenogram in ABPA may show transient, recurrent infiltrates or may suggest the presence of proximal bronchiectasis. High-resolution chestCT is a sensitive, noninvasive technique for the recognition of proximal bronchiectasis. The bronchial asthma of ABPA likely involves an IgE-mediated hypersensitivity, whereas the bronchiectasis associated with this disorder is thought to result from a deposition of immune complexes in proximal airways. Treatment usually requires the long-term use of systemic glucocorticoids.

Tropical eosinophilia is usually caused by filarial infection; however, eosinophilic pneumonias also occur with other parasites such as *Ascaris*, *Ancyclostoma* sp., *Toxocara* sp., and *Strongyloides stercoralis*. Tropical eosinophilia due to *Wuchereria bancrofti* or *W. malayi* occurs most commonly in southern Asia, Africa, and South America, and is treated successfully with diethylcarbamazine.

Drug-induced eosinophilic pneumonias are exemplified by acute reactions to nitrofurantoin, which may begin 2 h to 10 days after nitrofurantoin is started, with symptoms of dry cough, fever, chills, and dyspnea; an eosinophilic pleural effusion accompanying patchy or diffuse pulmonary infiltrates may also occur. Other drugs associated with eosinophilic pneumonias include sulfonamides, penicillin, chlorpropamide, thiazides, tricyclic antidepressants, hydralazine, mephenesin, mecamylamine, nickel carbonyl vapor, gold salts, isoniazid, para-aminosalicylic acid, and others. Treatment consists of withdrawal of the incriminated drugs and the use of glucocorticoids, if necessary. The eosinophilia-myalgia syndrome, caused by dietary supplements of L-tryptophan, is occasionally associated with pulmonary infiltrates.

The group of idiopathic eosinophilic pneumonias consists of diseases of varying severity. *Loeffler's syndrome* was originally reported as a benign, acute eosinophilic pneumonia of unknown cause characterized by migrating pulmonary infiltrates and minimal clinical manifestations. In some patients, these clinical characteristics may prove to be secondary to parasites or drugs. *Acute eosinophilic pneumonia* has been

described recently as an idiopathic acute febrile illness lasting less than 7 days with severe hypoxemia, pulmonary infiltrates, and no history of asthma. *Chronic eosinophilic pneumonia* presents with significant systemic symptoms including fever, chills, night sweats, cough, anorexia, and weight loss lasting for several weeks to months. The chest x-ray classically shows peripheral infiltrates resembling a photographic negative of pulmonary edema. Some patients also have bronchial asthma of the intrinsic or nonallergic type. Dramatic clearing of symptoms and chest x-rays is often noted within 48 h after initiation of glucocorticoid therapy.

Allergic angiitis and granulomatosis of Churg and Strauss is a multisystem vasculitic disorder that frequently involves the skin, kidney, and nervous system in addition to the lung. The disorder may occur at any age and favors persons with a history of bronchial asthma. The asthma often is progressive until the onset of fever and exaggerated eosinophilia, at which time the symptoms of asthma may ease. The illness may be fulminating and the prognosis grave unless treated aggressively with glucocorticoids and, at times, immunosuppressive therapy. The recent introduction of leukotriene-modifying agents (zafirlukast, zyleuton, and montelukast) has unmasked a number of cases of unrecognized Churg-Strauss syndrome when individuals with asthma have been weaned from glucocorticoids with the use of these antigens.

The *hypereosinophilic syndrome* is characterized by the presence of more than 1500 eosinophils per microliter of peripheral blood for 6 months or longer; lack of evidence for parasitic, allergic, or other known causes of eosinophilia; and signs or symptoms of multisystem organ dysfunction. Consistent features are blood and bone marrow eosinophilia with tissue infiltration by relatively mature eosinophils. The heart may be involved with tricuspid valve abnormalities or endomyocardial fibrosis and a restrictive, biventricular cardiomyopathy. Other organs affected typically include the lungs, liver, spleen, skin, and nervous system. Treatment consists of glucocorticoids and/or hydroxyurea, plus treatment as needed for cardiac dysfunction, which is frequently responsible for much of the morbidity and mortality in this syndrome.

(Bibliography omitted in Palm version)

Back to Table of Contents

254. ENVIRONMENTAL LUNG DISEASES - *Frank E. Speizer*

This**chapter**provides perspectives on ways to assess pulmonary diseases for which environmental causes are suspected. This assessment is important because removal of the patient from a harmful environment is often the only intervention that might prevent further significant deterioration or lead to improvement in a patient's condition. Furthermore, the identification of an environment-associated disease in a single patient may lead to primary preventive strategies affecting other similarly exposed people who have not yet developed disease.

The exact magnitude of the problem is unknown, but there is no question that large numbers of people are at risk for developing serious respiratory disease as a result of occupational or environmental exposures. For example, recent estimates suggest that approximately 2.4 million workers in the United States have been exposed to crystalline silica or asbestos dust in mining and nonmining industries. Even if only 5% of these workers (a conservative estimate) are to suffer from respiratory disease as a result of their exposure, this figure represents more than 100,000 individuals.

Although industries are required to spend substantial amounts of capital in efforts to protect their workers, occupationally related respiratory diseases continue to occur. These diseases are often attributed to exposures in the past, at a time when we were not aware of the risk incurred and the need for worker protection to the degree that we are today.

HISTORY AND PHYSICAL EXAMINATION

The patient's history is of paramount importance in assessing any potential occupational or environmental exposure, and the physician must ask the patient to describe a suspected environmental exposure in detail.

Inquiry into specific work practices should include questions about specific contaminants involved, the availability and use of personal respiratory protection devices, the size and ventilation of workspaces, and whether coworkers have similar complaints. In addition, the patient must be questioned about alternative sources for potentially toxic exposures, including hobbies or other environmental exposures at home. Short-term exposures to potential toxic agents in the distant past must also be considered (Chap. 391).

Many people are aware of the potential hazards in their workplaces, and many states require that employees be informed about potentially hazardous exposures. These requirements include the provision of specific educational materials (including Material Safety Data Sheets), personal protective equipment and instructions in its use, and information on environmental control procedures. Reminders posted in the workplace may warn workers about hazardous substances. Protective clothing, lockers, and shower facilities may be considered necessary parts of the job. However, even in these more progressive industries, the introduction of new processes, particularly when related to the use of new chemical compounds, may change exposure significantly, and often only the employee on the production line is aware of the change. For the physician who regularly sees patients from a particular industry, a visit to the work site can be very instructive. Alternatively, physicians can request inspections by appropriate federal

and/or state authorities.

The physical examination of patients with environment-related lung diseases may help to determine the nature and severity of the pulmonary condition. Unfortunately, the pulmonary response to most injurious agents is the development of a limited number of nonspecific physical signs. These findings do not point to the specific causative agent, and other types of information must be used to arrive at an etiologic diagnosis.

PULMONARY FUNCTION TESTS AND CHEST RADIOGRAPHY

Many mineral dusts produce characteristic alterations in the mechanics of breathing and lung volumes that clearly indicate a restrictive pattern (Chaps. 250 and259). Exposures to a number of organic dusts or chemical agents capable of producing occupational asthma result in pronounced obstructive patterns of pulmonary dysfunction that may be reversible (Chap. 252). Measurement of change in forced expiratory volume (FEV_1) before and after a working shift can be used to detect an acute inflammatory or bronchoconstrictive response. An acute decrement of FEV_1over the first work shift of the week is a characteristic feature of cotton textile workers with byssinosis.

The chest radiograph is useful in detecting and monitoring the pulmonary response to mineral dusts. The International Labour Organization (ILO) International Classification of Radiographs of Pneumoconioses classifies chest radiographs according to the nature and size of opacities seen and the extent of involvement of the parenchyma. In general, opacities may be round or irregular, small (<10 mm in diameter) or large. They may be few in number, with visible normal lung markings, partially obscure normal markings, or totally obscure normal markings. Although useful for screening large numbers of workers, the ILO system lacks specificity and may over- or underestimate the functional impact of pneumoconiosis. With dusts causing rounded, regular opacities like those evident in coal worker's pneumoconiosis, the degree of involvement on the chest radiograph may be extensive, while pulmonary function may be only minimally impaired.

In contrast, in pneumoconiosis causing linear, irregular opacities like those seen in asbestosis, the radiograph may lead to underestimation of the severity of the impairment. It is possible for a patient to have a history of exposure, a moderately reduced forced vital capacity (FVC), and a reduced diffusion capacity in asbestosis with a relatively normal chest radiograph. The radiographic findings of irregular or linear opacities are simply more difficult to separate from normal markings until relatively late in the disease. When shadows become large, as shown in Fig. 254-1, the condition is termed *complicated pneumoconiosis*, sometimes called progressive massive fibrosis (PMF). For the individual patient with a history of exposure, conventional computed tomography (CT) and high-resolution computed tomography (HRCT) have improved the sensitivity of identifying diffuse parenchymal abnormalities of the lung. The procedures have been shown to provide earlier detection of silicosis and asbestosis.

Other diagnostic procedures of use in identifying environment-induced lung disease include evaluation of heavy metal concentrations in urine (arsenic in smelter workers, cadmium in battery plant workers); bacteriologic studies (tuberculosis in medical care personnel, anthrax in wool sorters); fungal studies (coccidioidomycosis in southwestern farm workers, histoplasmosis in poultry or pigeon handlers); and serologic studies

(psittacosis in pet shop workers or owners of sick birds, Q fever in tanners or slaughterhouse workers). Ultimately, a lung biopsy may be required both for morphologic diagnosis of the underlying pulmonary disease and for attempted identification of the specific etiologic agent.

MEASUREMENT OF EXPOSURE

If reliable environmental sampling data are available, this information should be used in assessing a patient's exposure. Since many of the chronic diseases result from exposure over many years, current environmental measurements should be combined with work histories to arrive at estimates of past exposure. Even in acute conditions, when monitoring of exposure may be possible, little may be known about the actual dose received by the lung. Most of the research on health effects of air pollutants (discussed later in this**chapter**) has relied on fixed-station monitoring of outdoor air, often at locations somewhat distant from the residences of the people being studied. In addition, most people spend less than 20% of their time outdoors. Therefore, outdoor measurements can be used only in a relative sense, and they cannot be relied on to estimate actual dose.

In situations where individual exposure to specific agents -- either in a work setting or via ambient air pollutants -- has been determined, transport of these agents through the airways may be an important factor affecting dose. Highly soluble gases such as sulfur dioxide are absorbed in the upper airway and presumably produce their effects by reflex response of sensitive neural fibrils in the trachea or larger airways. In contrast, nitrogen dioxide, which is less soluble, may reach the bronchioles and alveoli in sufficient quantities to result in an acute life-threatening disease in farmers exposed even briefly to the gas evolved from moldy hay in silos (silo-filler's disease).

Particle size and chemistry of air contaminants also must be considered. Particles above 10 to 15 um in diameter, because of their settling velocities in air, do not penetrate beyond the upper airways. These larger particles are often referred to as "fugitive dusts" and include pollens, other windblown dusts, and dusts resulting from mechanical industrial processes. They have little or no role in chronic respiratory disease except perhaps as related to cancer (see below).

Particles below 10 um in size are created by the burning of fossil fuels or high-temperature industrial processes resulting in condensation products from gases, fumes, or vapors. These particles are divided into two size fractions on the basis of their chemical characteristics. Particles of approximately 2.5 to 10 um (coarse-mode fraction) contain crustal elements, such as silica, aluminum, and iron. These particles mostly deposit relatively high in the tracheobronchial tree. Although the total mass of an ambient sample is dominated by these larger respirable particles, the number of particles, and therefore the surface area on which potential toxic agents can deposit and be carried to the lower airways, is dominated by particles smaller than 2.5 um (fine-mode fraction or accumulation mode). The smallest particles, those less than 0.1 um in size, remain in the airstream and deposit in the lung only on a random basis as they come into contact with the alveolar walls.

Besides the size characteristics of particles and the solubility of gases, the actual

chemical composition, mechanical properties, and immunogenicity or infectivity of inhaled material determine in large part the nature of the diseases found among exposed persons. Few studies to date have directly measured those characteristics. However, they are of increasing concern as management strategies for environmental and occupational exposures are developed.

OCCUPATIONAL EXPOSURES AND PULMONARY DISEASE

ASBESTOSIS

Except in localized regions with single industrial exposures, such as coal-mining or granite-quarrying regions, the most frequent inorganic dust-related chronic pulmonary diseases are associated with industries using *asbestiform fibers*. *Asbestos* is a generic term for several different mineral silicates, including chrysolite, amosite, anthophyllite, and crocidolite. Besides workers involved in the mining, milling, and manufacturing of asbestos products, workers in the building trades, including pipe fitters and boilermakers, were exposed to asbestos, which was widely used in construction because of its exceptional thermal and electric insulation properties. In addition, asbestos was used in the manufacture of fire-smothering blankets and safety garments, as filler for plastic materials, in cement and floor tiles, and in friction materials, such as brake and clutch linings.

Exposure to asbestos is not limited to persons who directly handle the material. Cases of asbestos-related diseases have been encountered in individuals with only moderate exposure, such as the painter or electrician who works alongside the insulation worker in a shipyard or the housewife who does no more than shake out and wash her husband's work clothes. Community exposure has probably resulted from the use of asbestos-containing material sprayed on steel girders in many large buildings as a safety feature to prevent buckling in case of fire.

Asbestos was first used extensively in the 1940s. Starting in 1975 it was mostly replaced with synthetic mineral fibers, such as fiberglass or slag wool. However, asbestos is still used in the manufacture of brake linings and remains as pipe and boiler insulation in hundreds of thousands of workplaces and homes. Despite current regulations mandating adequate training for any worker potentially exposed to asbestos, exposure probably continues among inexperienced demolition workers. The major health effects from exposure to asbestos are pulmonary fibrosis (asbestosis) and cancers of the respiratory tract, the pleura, and (in rare cases) the peritoneum.

Asbestosis is a diffuse interstitial fibrosing disease of the lung that is directly related to the intensity and duration of exposure. Except for its association with a history of exposure to asbestos (generally in a work setting), asbestosis resembles the other forms of diffuse interstitial fibrosis (Chap. 259). Usually, moderate to severe exposure has taken place for at least 10 years before the disease becomes manifest.

Physiologic studies reveal a restrictive pattern with a decrease in lung volumes. Flow rates are commonly reduced less than would be predicted on the basis of the volume reduction. An early sign of severe disease may be a reduction in diffusing capacity.

Pulmonary fibrosis may occur following sufficient exposure to any of the asbestiform fiber types. The fibrotic lesions do not appear to relate to either shape or chemical composition of any fiber type. During phagocytosis of the asbestos fiber, the membrane of the macrophage is damaged and this damage results in the release of lysosomes containing enzymes that may act to damage the lung parenchyma. The clinical manifestations are typical of those physical findings in any patient with pulmonary fibrosis (Chap. 259).

Diagnosis The chest radiograph can be used to detect a number of manifestations of asbestos exposure as well as to identify specific lesions. Past exposure is specifically indicated by pleural plaques, which are characterized by either thickening or calcification along the parietal pleura, particularly along the lower lung fields, the diaphragm, and the cardiac border. Without additional manifestations, pleural plaques imply only exposure, not pulmonary impairment. Benign pleural effusions may occur, particularly in patients with asbestosis, but are not necessarily restricted to those with overt disease. The fluid is sterile but may be a serous or blood-stained exudate and may occur bilaterally. The effusion may be slowly progressive or may resolve spontaneously.

The radiographic diagnosis of asbestosis depends on the presence of irregular or linear opacities, usually first noted in the lower lung fields and spreading into the middle and upper lung fields as the disease progresses. An indistinct heart border or a "ground glass" appearance in the lung fields is seen in some cases. As the fibrotic changes in the parenchyma begin to coalesce, the patient develops obliteration of entire acinar units, with eventual formation of the classical honeycombed lung, which appears on chest radiographs as coarse infiltrates with small (about 7- to 10-um) air spaces. In cases in which the x-ray changes are less obvious,HRCT may show distinct changes of subpleural curvilinear lines 5 to 10 cm in length that appear to be parallel to the pleural surface; these alterations increase the positive predictive value of radiographic evidence from approximately 85% to about 100%.

In general, newly diagnosed cases will have resulted from exposure levels that were present many years before and, in spite of the patients' having left the industry, are attributable to that former exposure. Since the patient may be eligible for compensation within a specific time frame after the diagnosis of an asbestos-related disease is made, the physician making the diagnosis should be certain to inform the patient promptly. On occasion, the physician may have reason to suspect ongoing exposure from a patient's current job description or actual monitoring data. In such cases, federal or state health authorities may need to be notified. Present-day occupational safety and health regulations, if followed properly, protect workers from exposure.

Casual, nonoccupational exposure to undisturbed sources of asbestos-containing materials -- e.g., in walls of schools or other buildings -- represents little if any hazard to people who inhabit or work in such buildings. Because the association of smoking and asbestos exposure increases the risk of developing lung cancer (see below), it is extremely important to advise patients with a history of exposure to asbestos to stop smoking. No specific therapy is available in the management of patients with asbestosis. The supportive care is the same as that given to any patient with diffuse interstitial fibrosis from any cause.

Lung cancer (Chap. 88), either squamous cell carcinoma or adenocarcinoma, is the most frequent cancer associated with asbestos exposure. The excess frequency of lung cancer in asbestos workers is associated with a minimum lapse of 15 to 19 years between first exposure and development of the disease. Persons with more exposure are at greater risk of disease. In addition, there appears to be a significant multiplicative effect that leads to a far greater risk of lung cancer in persons who are cigarette smokers and have asbestos exposure than would be expected from the additive risk of each factor. To date, efforts to consider these high-risk individuals for special surveillance studies, including sputum cytologic examinations and repeated chest x-rays as frequently as every 4 to 6 months, have resulted in neither significant early detection nor prolonged survival once the lung cancer is found.

Mesotheliomas (Chap. 262), both pleural and peritoneal, are also associated with asbestos exposure. In contrast to lung cancers, these tumors do not appear to be associated with smoking. Relatively short-term asbestos exposures of 1 to 2 years or less occurring some 20 to 25 years in the past have been associated with the development of mesotheliomas (an observation that emphasizes the importance of obtaining a complete environmental exposure history). The risk for this type of tumor peaks 30 to 35 years after initial exposure. Since maximum exposure took place in the United States between 1930 and 1960, peak incidence of disease in men occurred in 1997, with a total of 2300 cases. Incidence is expected to decline over the next 30+ years to about 500 cases per year.

Although approximately 50% of mesotheliomas metastasize, the tumor generally is locally invasive, and death usually results from local extension. Most patients present with effusions that may obscure the underlying pleural tumor. In contrast to the findings in effusion due to other causes, because of the restriction placed on the chest wall, no shift of mediastinal structures toward the opposite side of the chest will be seen. The major diagnostic problem is differentiation from peripherally spreading pulmonary adenocarcinoma or from adenocarcinoma metastasized to pleura from an extrathoracic primary site. Although a needle biopsy may be diagnostic, an open biopsy is often necessary, and even the latter procedure may not provide a definitive diagnosis of the origin of the tumor.

Since epidemiologic studies have shown that more than 80% of mesotheliomas may be associated with asbestos exposure, documented mesothelioma in a worker with occupational exposure to asbestos may be compensable in many parts of the United States.

SILICOSIS

In spite of the technical adequacy of existing protective equipment, *free silica* (SiO_2), or crystalline quartz, is still a major occupational hazard. In the United States, estimates of potential numbers of exposed workers range between 1.2 and 3 million people. The major occupational exposures include: mining; stonecutting; employment in abrasive industries, such as stone, clay, glass, and cement manufacturing; foundry work; packing of silica flour; and quarrying, particularly of granite. Most often, progressive pulmonary fibrosis (silicosis) occurs in a dose-response fashion after many years of exposure.

Workers exposed through sandblasting in confined spaces, tunneling through rock with high quartz content (15 to 25%), or the manufacture of abrasive soaps may develop acute silicosis with as little as 10 months' exposure. The disease may be rapidly fatal in less than 2 years, despite the discontinuation of exposure. A radiographic picture of profuse miliary infiltration or consolidation is characteristic of acute silicosis.

In long-term, less intense exposure, small rounded opacities in the upper lobes, with retraction and hilar adenopathy, classically appear on the radiograph after 15 to 20 years. Calcification of hilar nodes may occur in as many as 20% of cases and produces the characteristic "eggshell" pattern. These changes may be preceded by or associated with a reticular pattern of irregular densities that are uniformly present throughout the upper lung zones.

The nodular fibrosis may be progressive in the absence of further exposure, with coalescence and formation of nonsegmental conglomerates of irregular masses in excess of 1 cm in diameter. These masses become quite large and are characteristic ofPMF(Figure 254-1). Significant functional impairment with both restrictive and obstructive components may be associated with this form of silicosis. In the late stages of the disease, ventilatory failure may develop. In more subtle cases,CT may be helpful both in identifying nodules, which are preferentially located in the posterior aspect of the upper lobes, as well as in identifying larger opacities and more coalescence than might be noted on regular chest x-rays. Patients with silicosis are at greater risk of acquiring *Mycobacterium tuberculosis* infections (silicotuberculosis) and atypical mycobacterial infections. Because the frequency with which tuberculosis has been found at autopsy in patients with PMF exceeds considerably the frequency of premorbid diagnosis, treatment for tuberculosis is indicated in any patient with silicosis and a positive tuberculin test.

Other less hazardous silicates include Fuller's earth, kaolin, mica, diatomaceous earths, silica gel, soapstone, carbonate dusts, and cement dusts. The production of fibrosis in workers exposed to these agents is believed to be related either to the free silica content of these dusts or, for substances that contain no free silica, to the potentially large dust loads to which these workers may be exposed.

Other silicates, including *talc dusts*, may be contaminated with asbestos and/or free silica. Accidental exposure to significant quantities of talc may result in an acute syndrome with cough, cyanosis, and labored breathing (acute talcosis). Severe progressive fibrosis with respiratory failure may ensue within a few years. Far more common is the fibrosis and/or pleural or lung cancer associated with chronic exposure in rubber workers who use commercial talc as a lubricant in tire molds. Pure talc does not produce fibrosis; thus, it is difficult to sort out whether the effects are due to the contamination of commercial talc by asbestos or by free silica.

COAL WORKER'S PNEUMOCONIOSIS (CWP)

Coal dust is associated with CWP, which has enormous social, economic, and medical significance in every nation in which coal mining is an important industry. Simple radiographically identified CWP is seen in 12% of all miners and in as many as 50% of anthracite miners with more than 20 years' work on the coal face. The prevalence of

disease is lower in workers in bituminous coal mines. Since much western U.S. coal is bituminous, CWP is less prevalent in that region.

Much of the symptomatology associated with simpleCWPappears to be similar and additive to the effects of cigarette smoking on the development of chronic bronchitis and obstructive lung disease (Chap. 258). In the early stages of simple CWP, radiographic abnormalities consist of small, irregular opacities (reticular pattern). With prolonged exposure, one sees small, rounded, regular opacities, 1 to 5 mm in diameter (nodular pattern). Calcification is generally not seen, although approximately 10% of older anthracite miners have calcified nodules.

ComplicatedCWP is manifested by the appearance on the chest radiograph of nodules ranging from 1 cm in diameter to the size of an entire lobe, generally confined to the upper half of the lungs. This condition, considered a form ofPMF, is accompanied by a significant reduction in diffusing capacity and is associated with premature mortality. In contrast to patients with silicosis, underground miners with simple CWP develop PMF at a rate of only 5 to 15%, depending on the type of coal.

The mechanism whereby PMFoccurs inCWP is not fully understood. Several hypotheses have been proposed, including: (1) sufficient free silica is present in the dust; (2) normal clearance mechanisms are unable to clear the excessive dust loads; and (3) atypical reactions to *M. tuberculosis* occur. As previously described, PMF in silicosis is associated with prolonged duration and high intensity of exposure to free silica. Heavy exposure to carbon particles free of silica occurs in carbon black, graphite, and charcoal workers. The prolonged exposure of these workers may result in sufficient accumulation of carbon in the lung to produce PMF. The mechanism appears to relate to a breakdown of the clearance capacity of the airways.

Caplan's syndrome (Chap. 312), first described in coal miners but subsequently found in patients with a variety of pneumoconioses, includes seropositive rheumatoid arthritis with characteristicPMF. The syndrome suggests an immunopathologic mechanism. Over the last decade, the mechanisms by which the chronic inhalation of mineral dusts produce an increase in inflammatory cells (including macrophages and neutrophils), which in turn causes PMF, have been explored. Coal dust can: (1) be a source of reactive oxygen species causing lung injury; (2) result in stimulation of macrophages to produce cytokines and enhance production of (anti)fibrogenic factors such as TNF-a; (3) increase protease activity; and (4) increase inactivation of a_1-antitrypsin and leukocyte elastase activity. The final pathologic pathway may be fibrosis resulting from the interactions of a variety of these mechanisms.

BERYLLIOSIS

Beryllium may produce an acute pneumonitis or, far more commonly, a chronic interstitial pneumonitis. Unless one inquires specifically about occupational exposures to beryllium in the manufacture of alloys, ceramics, high-technology electronics, and (before the 1950s) the production of fluorescent lights, one may miss entirely the etiologic relationship to an occupational exposure. Nonspecific pulmonary function tests may be normal or may indicate evidence of restrictive disease. Between 2 and 15 years of exposure, depending on its intensity, are required for the disease to become

manifest. On open lung biopsy, granulomatous formation similar to that seen in sarcoidosis (Chap. 318) may make differentiation impossible unless tissue levels of beryllium are measured.

Other hard metals, including aluminum powders, chromium, cobalt, titanium dioxide, and tungsten, may produce an interstitial pneumonitis, but this is rare.

OTHER INORGANIC DUSTS

Other dusts are considered *nuisance dusts* because their major environmental and health effects seem to be reduction in visibility and irritation of eyes, ears, nasal passages, and other mucous membranes, respectively. If they penetrate to the lower airways, these dusts do not affect the architecture of the terminal bronchioles or acinar spaces nor do they destroy collagen. Generally, clinical effects are reversible. Pulmonary function tests are usually normal unless another disease process coexists. If the dusts are radiodense, macular collections may produce striking radiographic pictures that are so characteristic that patients with a history of significant exposure are easily diagnosed as having the condition that bears the name reflecting the nature of the dust. Examples of radiodense dusts include iron and iron oxides from welding or silver finishing (*siderosis*); tin oxide used in metallurgy, color stabilization, printing, and the manufacture of porcelain, glass, and fabric (*stannosis*); and barium sulfate used as a catalyst for organic reactions, drilling mud components, and electroplating (*baritosis*). Other metal dusts producing similar radiodense pictures include *cerium dioxide* and *antimony salts*.

Most of the inorganic dusts discussed thus far are associated with the production of either dust macules or interstitial fibrotic changes in the lung. Another set of dusts (Table 254-1), along with some of the dusts previously discussed, is associated with chronic mucous hypersecretion (chronic bronchitis), with or without reduction of expiratory flow rates. These conditions are caused by cigarette smoking, and any effort to attribute some component of the disease to occupational and environmental exposures must take cigarette smoking into account. Most studies suggest an additive effect of dust exposure and smoking. The pattern of the effect is similar to that of cigarette smoking, suggesting that small airway inflammation may be the initial site of pathologic response in those cases associated with the development of obstructive lung disease. Cigarette smoke is usually the more noxious agent, and dust effects may be discernible only in nonsmokers.

ORGANIC DUSTS

Some of the specific diseases associated with organic dusts are discussed in detail in the chapters on asthma (Chap. 252) and hypersensitivity pneumonitis (Chap. 253). Many of these diseases are named for the specific setting in which they are found, e.g., farmer's lung, malt worker's disease, or mushroom worker's disease. Occupational and other environmental exposures must be sought when these conditions are suspected. Often the temporal relation of symptoms to exposure furnishes the best evidence for the diagnosis. Three occupational groups are singled out for discussion because they represent the largest proportion of people affected by the diseases resulting from organic dusts.

Cotton Dust (Byssinosis) Estimates of the number of exposed persons in the United States vary, but probably over 800,000 persons are exposed occupationally to cotton, flax, or hemp in the production of yarns for cotton, linen, and rope making. Although this discussion focuses on cotton, the same syndrome -- albeit somewhat less severe -- has been reported in association with exposure to flax, hemp, and jute.

Exposure occurs throughout the manufacturing process but is most pronounced in those portions of the factory involved with the treatment of the cotton prior to spinning -- i.e., blowing, mixing, and carding (straightening of fibers). Attempts to control dust levels by use of exhaust hoods, general increases in ventilation, and wetting procedures in some settings have been highly successful. However, respiratory protective equipment appears to be required during certain operations to prevent workers from being exposed to levels of dust that exceed the current U.S. cotton dust standard.

Byssinosis is characterized clinically as occasional (early stage) and then regular (late stage) chest tightness toward the end of the first day of the workweek ("Monday chest tightness"). In epidemiologic studies, depending on the level of exposure via the carding room air, up to 80% of employees may show a significant drop in theirFEV_1over the course of a Monday shift.

Initially the symptoms do not recur on subsequent days of the week. However, in 10 to 25% of workers, the disease may be progressive, with chest tightness recurring or persisting throughout the workweek. After more than 10 years of exposure, workers with recurrent symptoms are more likely to have an obstructive pattern on pulmonary function testing. These higher grades of impairment are seen in workers exposed both to high levels of dust and for greater durations. There is an additive effect of cotton dust exposure plus cigarette smoking. The highest grades of impairment are generally seen in smokers.

Treatment in the early stages of the disease is directed toward reversing the bronchospasm with bronchodilators; however, the chest tightness appears to relate, at least in part, to histamine release, and antihistamines have been shown to lessen the anticipated fall in$FEV_1$the first day of the week. Clearly, reduction of dust exposure is of primary importance. All workers with persistent symptoms or significantly reduced levels of pulmonary function should be moved to areas of lower risk of exposure. Regular surveillance of pulmonary function in the industry has made it easier to identify affected persons. Persons with reduced pulmonary function, a personal history of respiratory allergy, and a history of continued cigarette smoking should be considered at increased risk of developing byssinosis in association with work in the cotton industry.

Grain Dust Although the exact number of workers at risk in the United States is not known, at least 500,000 people work in grain elevators, and over 2 million farmers are potentially exposed to grain dust. The presentation of disease in grain elevator employees or in workers in flour or feed mills is virtually identical to the characteristic findings in cigarette smokers, i.e., persistent cough, mucous hypersecretion, wheeze and dyspnea on exertion, and reduced$FEV_1$and FEV_1/FVCratio (Chap. 250).

Dust concentrations in grain elevators vary greatly but appear to be in excess of 10,000

ug/m3; approximately one-third of the particles, by weight, are in the respirable range. The effect of grain dust exposure is additive to that of cigarette smoking, with approximately 50% of workers who smoke having symptoms. Among nonsmoking grain elevator operators, approximately one-quarter have mucous hypersecretion, about five times the number that would be expected in unexposed nonsmokers. However, evidence of obstruction on pulmonary function studies is observed only in workers who smoke. It is not clear whether the reason is an enhancement of the cigarette smoking effect in exposed workers or a greater susceptibility of smokers to the effects of grain dust.

Farmer's Lung This condition results from exposure to moldy hay containing spores of thermophilic actinomycetes that produce a hypersensitivity pneumonitis (Chap. 253). There are few good population-based estimates of the frequency of occurrence of this condition in the United States. However, among farmers in Great Britain, the rate of disease ranges from approximately 10 to 50 per 1000. The prevalence of disease varies in association with rainfall, which determines the amount of fungal growth, and with differences in agricultural practices related to turning and stacking hay.

The patient with acute farmer's lung presents 4 to 8 h after exposure with fever, chills, malaise, cough, and dyspnea without wheezing. The history of exposure is obviously essential to distinguish this disease from influenza or pneumonia with similar symptoms. In the chronic form of the disease, the history of repeated attacks after similar exposure is important in differentiating this syndrome from other causes of patchy fibrosis (e.g., sarcoidosis).

A wide variety of other organic dusts are associated with the occurrence of hypersensitivity pneumonitis (Chap. 253). For those patients who present with hypersensitivity pneumonitis, specific and careful inquiry about occupations, hobbies, or other home environmental exposures will, in most cases, reveal the source of the etiologic agent.

ASSESSMENT OF DISABILITY

Significant reduction of dust levels in coal mines has resulted from federal legislation, enacted in the United States in 1969, that requires that respirable dust levels in underground mines be reduced to less than 2000 ug/m3. This same legislation authorizes payment to coal miners (or their survivors) totally disabled byCWP. The criteria for disability from CWP remain unclear and arbitrary. It is critical that physicians involved in occupational lung disease claim cases be aware of detailed exposure histories of their patients, in terms of both occupational exposures and other environmental exposures (cigarette smoking). To assess disability properly may require input not only from physicians but also from experts in ergonomics and vocational rehabilitation, lawyers, and employer and employee representatives.

Most commonly, the patient presents with asthma, and it is the physician's task to decide whether the asthma is occupation-induced or work-aggravated asthma. The distinction is important not only because of the implications for disability compensation but also because the longer one is exposed to an inciting agent, the worse the prognosis for recovery from occupation-induced asthma. The clinical evaluation of such

a patient requires adherence to a prescribed protocol that may include not only the components of the evaluation previously described but also rechallenge of the patient in a controlled setting or under a carefully monitored program in a work setting.

TOXIC CHEMICALS

Exposure to toxic chemicals affecting the lung generally involves gases and vapors. A common accident is one in which the victim is trapped in a confined space where the chemicals have accumulated to toxic levels. In addition to the specific toxic effects of the chemical, the victim will often sustain considerable anoxia, which can play a dominant role in determining whether the individual survives.

Table 254-2 lists a variety of toxic agents that can produce acute and sometimes life-threatening reactions in the lung. All these agents in sufficient concentrations have been demonstrated, at least in animal studies, to affect the lower airways and disrupt alveolar architecture, either acutely or as a result of chronic exposure. Some of these agents may be generated acutely in the environment. For example, when plastics burn, a number of compounds, including hydrogen cyanide and hydrochloric acid, may be formed and released.**The effects and treatment of exposure to these toxic gases are discussed in Chap. 391.*

Firefighters and fire victims are at risk of *smoke inhalation*, a numerically important cause of acute cardiorespiratory failure. Smoke inhalation kills more fire victims than does thermal injury. Carbon monoxide poisoning with resulting significant hypoxemia can be life-threatening (Chap. 396). Firefighters may inappropriately use the "blackness" of the smoke to indicate the degree of incomplete combustion and thus of carbon monoxide elevation. The use of synthetic materials (plastic, polyurethanes), which, when burned, may release a variety of other toxic agents, must be considered when evaluating smoke inhalation victims. Exposed victims may suffer some degree of lower respiratory tract inflammation, similar to that seen with exposure to other irritant gases (e.g., chlorine). Severe cases may include pulmonary edema.

Firefighters and victims also may be exposed to large quantities of particulate smoke. Significant long-term effects are not clearly associated with this particulate exposure except as related to the production of irritating effects on the upper airways; however, increased airway responsiveness in firefighters with repeated episodes of smoke inhalation has been demonstrated.

Some agents used in the manufacture of synthetic materials such as plastics, polyurethanes, and other polymers have resulted in some workers' being sensitized to extremely low levels of *isocyanates*, *aromatic amines*, or *aldehydes*. Repeated exposure to these agents causes some workers to develop chronic cough and sputum production, asthma, or episodes of low-grade fever and malaise.

Exposure occurs by an unusual route in *polymer fume fever*. Polymers, notably fluorocarbons, which at normal temperatures produce no reaction, may be transmitted from a worker's hands to his or her cigarettes. As the cigarette burns, the polymer is volatilized, and the inhaled agent causes a characteristic syndrome of fever, chills, malaise, and occasionally mild wheezing. The same scenario applies when workers are

exposed to heated polymers without cigarette use -- *meat wrappers' asthma*. A similar self-limited, influenza-like syndrome -- *metal fume fever* -- results from acute exposure to fumes or smoke of zinc, copper, magnesium, and other volatilized metals. The syndrome may begin several hours after work and resolves within 24 h, only to return on repeated exposure. A proper occupational history should make the diagnosis evident.

ENVIRONMENTAL RESPIRATORY CARCINOGENS

Historically, it has been the astute clinician who has recognized a higher incidence of malignant tumors associated with certain environmental exposures. When these observations are linked to an occupational setting, they must be pursued by epidemiologic studies of relatively large groups of both current and former workers. Often the concentration and/or exact nature of the substances involved in the putative exposures cannot be determined. Rarely, the possibility that a substance can play an etiologic role in cancer is supported by observing that a few cases of a very rare tumor in a particular group represent "an epidemic." Examples are nasal sinus and lung cancer in nickel workers, angiosarcomas of the liver in vinyl chloride workers, and adenocarcinomas of the nose in woodworkers.

Only in those few cases in which animal studies have been carried out can one confirm that a given suspected agent is really a carcinogen. For example, bis(chloromethyl)ether (BCME) has been shown to produce tumors in animals and oat cell cancer of the lung in humans. In this particular case, BCME, used as a chemical intermediary in the manufacture of a number of organic compounds, was found to produce tumors in animals at about the same time as the substance was introduced into industry.

In addition to asbestos exposures, other occupational exposures associated with either proven or suspected respiratory carcinogens include those to acrylonitrile, arsenic compounds, beryllium (animal studies only),BCME, chromium, polycyclic hydrocarbons (through coke oven emissions), iron oxide, isopropyl oil (nasal sinuses), mustard gas, the various ores used to produce pure nickel, talc (possible asbestos contamination in both mining and milling), vinyl chloride, welding materials, wood used in woodworking (nasal cancer only), and uranium. The occurrence of excess cancers in uranium miners raises the possibility that a large number of workers are at risk by virtue of exposure to similar radiation hazards. This number includes not only workers involved in processing uranium but also workers exposed in underground mining operations where radon daughters may be emitted from rock formations.

GENERAL ENVIRONMENTAL EXPOSURES

AIR POLLUTION

Dramatic and disastrous episodes of air pollution inversion have been documented in many industrialized centers in the world. Each of these episodes has been associated with excess acute mortality in the very old, the very young, and those with chronic cardiopulmonary diseases. The most dramatic event was the London fog of 1952, in which approximately 4000 excess deaths occurred over a 2-week period following 5 days of severe cold and dense fog. Similar episodes in the United States, although less dramatic in terms of total deaths, occurred in Donora, Pennsylvania, in 1948 and in New

York City in the 1960s. In these episodes, which were generally associated with cold temperature and air stagnation, patients with underlying cardiopulmonary disease were most severely affected.

In addition to significant excess mortality during these episodes, a large number of people required medical care for cardiorespiratory complaints. Subsequent follow-up studies failed to implicate these episodic disasters in the etiology of chronic respiratory disease in adults. On the other hand, many epidemiologic studies of both international and regional differences in the prevalences of chronic respiratory disease suggest that long-term exposures in polluted areas in the early to middle part of the twentieth century were associated with excess chronic respiratory disease.

In 1970, the U.S. government established air quality standards for several pollutants believed to be responsible for excess cardiorespiratory diseases. Primary standards regulated by the Environmental Protection Agency (EPA) designed to protect the public health with an adequate margin of safety exist for sulfur dioxide, particulates <10 um in size, nitrogen dioxide, ozone, lead, and carbon monoxide. Standards for each of these pollutants are updated regularly through an extensive review process conducted by the EPA. In 1997, a new standard was added for particles less than 2.5 um; however, the standard does not become effective until year 2002.

Pollutants are generated from both stationary sources (power plants and industrial complexes) and mobile sources (automobiles), and none of the pollutants occurs in isolation. Thus, except for the change in carboxyhemoglobin from carbon monoxide exposure, it becomes extremely difficult to relate any specific health effect to any single pollutant. Furthermore, pollutants may be changed by chemical reactions after being emitted. For example, reducing agents, such as sulfur dioxide and particulate matter from a power plant stack, may react in air to produce acid sulfates and aerosols, the precursors of acid rain, which can be transported long distances in the atmosphere. Oxidizing substances, such as oxides of nitrogen and oxidants from automobile exhaust, may react with sunlight to produce ozone. Although originally a problem confined to the southwestern part of the United States, in recent years, at least during the summertime, elevated ozone and acid aerosol levels have been documented throughout the United States. Both acute and chronic effects of these exposures are currently under investigation.

The symptoms and diseases associated with air pollution are the same as the nononcogenic conditions commonly associated with cigarette smoking. In addition, respiratory illness in early childhood has been associated with chronic exposure to only modestly elevated levels of SO_2 and respirable particles. Recent population-based studies comparing cities that have relatively high levels of particulate exposures with less polluted communities suggest excess morbidity and mortality from cardiorespiratory conditions in long-term residents of the former communities. This finding, in part, has led to greater emphasis on publicizing pollution alert levels. One can only advise individuals with significant cardiopulmonary impairment to stay indoors during periods when pollution exceeds current standards.

INDOOR EXPOSURE

Because of increased concern about energy costs, efforts to become energy efficient have led to reduced air-exchange rates in indoor environments. The unintentional effect of these efforts has been to increase exposures to a variety of air contaminants heretofore not considered important.

Until relatively recently, little attention was given to the effects of *passive cigarette smoking* (Chap. 390). Several studies have shown that the respirable particulate load in any household is directly proportional to the number of cigarette smokers living in the home. Increases in prevalence of respiratory illnesses and reduced levels of pulmonary function measured with simple spirometry have been found in children of smoking parents in a number of studies.

Evidence from numerous case-control and cohort studies shows modest excess disease associations for cardiopulmonary diseases and lung cancer. Because most of these excess relative risks appear to be below 50%, it is virtually impossible for any one of the studies to be considered definitive. Thus, the techniques of meta-analysis have been used effectively to combine data from the best of these studies. The most recent meta-analyses for lung cancer, cardiac disease, and respiratory disease in terms of excess mortality suggest an approximately 25% increase for each condition, even after adjustment for major potential confounders. According to measures of plasma cotinine, a metabolite of nicotine, a nonsmoker living with a smoker is exposed to approximately 1% of the level of tobacco smoke to which a smoker of 20 cigarettes a day is exposed. In spite of some prominent detractors, these combined relative risks appear to be consistent with the estimated exposure levels and suggest a consensus that the associations are causal.

Radon gas is believed to be a risk factor for lung cancer. The main radon product (radon 222) is a gas that results from the decay series of uranium 238, with the immediate precursor being radium 226. The amount of radium in earth materials determines how much radon gas will be emitted. Outdoors, the concentrations are trivial. Indoors, levels are dependent on the ventilation rate and the size of the space into which the gas is emitted. Levels associated with excess lung cancer risk may be present in as many as 10% of the houses in the United States. When smokers reside in the household, the problem is potentially greater, since the molecular size of radon particles allows them to readily attach to smoke particles that are inhaled. Fortunately, technology is available for assessing and reducing the level of exposure.

Other indoor exposures associated with an increased risk of atopy and asthma include those to such specific recognized putative biologic agents as cockroach antigen, dust mites, and pet danders. Other indoor chemical agents include formaldehyde, perfumes, and latex particles. Of recent interest are the nonspecific responses associated with "tight-building syndrome," in which no particular agent has been implicated; the affected individuals suffer from a wide variety of complaints, including respiratory symptoms, that are relieved only by avoiding exposure in the building in question. The degree to which "smells" or other sensory stimuli are involved in the triggering of potentially incapacitating psychological or physical responses has yet to be determined, and the long-term consequences of such environmental exposures are as yet unknown.

PORTAL OF ENTRY

The lung is a primary point of entry into the body for a number of toxic agents that affect other organ systems. For example, the lung is a route of entry for benzene (bone marrow), carbon disulfide (cardiovascular and nervous systems), cadmium (kidney), and metallic mercury (kidney, central nervous system). Thus, in any disease state of obscure origin, it is important to consider the possibility of inhaled environmental agents. Such consideration can sometimes furnish the clue needed to identify a specific external cause for a disorder that might otherwise be labeled "idiopathic."

(Bibliography omitted in Palm version)

Back to Table of Contents

255. PNEUMONIA, INCLUDING NECROTIZING PULMONARY INFECTIONS (LUNG ABSCESS) - *Matthew E. Levison*

Pneumonia is an infection of the pulmonary parenchyma that can be caused by various bacterial species, including mycoplasmas, chlamydiae, and rickettsiae; viruses; fungi; and parasites (Table 255-1). Thus pneumonia is not a single disease but a group of specific infections, each with a different epidemiology, pathogenesis, clinical presentation, and clinical course. Identification of the etiologic microorganism is of primary importance, since this is the key to appropriate antimicrobial therapy. However, because of the serious nature of the infection, antimicrobial therapy generally needs to be started immediately, often before laboratory confirmation of the causative agent. The specific microbial etiology remains elusive in more than one-third of cases -- e.g., when no sputum is available for examination, blood cultures are sterile, and there is no pleural fluid. Serologic confirmation requires weeks because of the late formation of specific antibody.

Thus initial antimicrobial therapy is often empirical and is based on the setting in which the infection was acquired, the clinical presentation, patterns of abnormality on chest radiography, results of staining of sputum or other infected body fluids, and current patterns of susceptibility of the suspected pathogens to antimicrobial agents. After the etiologic agent is identified, specific antimicrobial therapy can be chosen.

DEFENSE MECHANISMS

The lung is a complex structure composed of aggregates of units that are formed by the progressive branching of the airways. Approximately 80% of the cells lining the central airways are ciliated, pseudostratified, columnar epithelial cells; the percentage decreases in the peripheral airways. Each ciliated cell contains about 200 cilia that beat in coordinated waves ~1000 times per minute, with a fast forward stroke and a slower backward recovery. Ciliary motion is also coordinated between adjacent cells so that each wave is propagated toward the oropharynx. The cilia are covered by a liquid film that is ~5 to 10 um thick and is composed of two layers. The outer, or gel, layer is viscous and traps deposited particles. The cilia beat in the less viscous inner, or sol, layer. During the forward stroke, the tips of the cilia just touch the viscous gel and propel it toward the oropharynx. During recovery, the cilia move entirely within the low-resistance sol layer. Ciliated cells are interspersed with mucus-secreting cells in the trachea and bronchi but not in the bronchioles.

The alveolar walls, from blood to air, consist of the endothelium that lines the network of anastomotic capillaries, the capillary basement membrane, the interstitial tissue, the alveolar basement membrane, the alveolar lining epithelial cells (which are either flattened type I pneumocytes that cover 95% of the alveolar surface or rounded, granular, surfactant-producing type II pneumocytes), and epithelial lining fluid. The epithelial lining fluid contains surfactant, fibronectin, and immunoglobulin, which may opsonize or -- in the presence of complement -- lyse microbial pathogens deposited on the alveolar surface. Loosely attached to the lining cells or lying free within the lumen are the alveolar macrophages, lymphocytes, and a few polymorphonuclear leukocytes.

The lower respiratory tract is normally sterile, despite being adjacent to enormous

numbers of microorganisms that reside in the oropharynx and being exposed to environmental microorganisms in inhaled air. This sterility is the result of efficient filtering and clearance mechanisms.

Infectious particles deposited on the squamous epithelium of distal nasal surfaces normally are removed by sneezing, while those deposited on the more proximal ciliated surfaces are swept posteriorly in the mucus lining into the nasopharynx, where they are swallowed or expectorated. Reflex closure of the glottis and cough protect the lower respiratory tract. Those particles deposited on the tracheobronchial surface are swept by ciliary motion toward the oropharynx. Infectious particles that bypass defenses in the airways and are deposited on the alveolar surface are cleared by phagocytic cells and humoral factors. Alveolar macrophages are the major phagocytes in the lower respiratory tract. Some phagocytosed microorganisms are killed by the phagocyte's oxygen-dependent systems, lysosomal enzymes, and cationic proteins. Other microorganisms can evade microbicidal mechanisms and persist within the macrophage. For example, *Mycobacterium tuberculosis* persists within the lysosome, while *Legionella* resides within intracellular inclusions that fail to fuse with lysosomes. Intracellular pathogens can then be transported to the ciliated surfaces and into the oropharynx or via the lymphatics to regional lymph nodes. The alveolar macrophages process and present microbial antigens to the lymphocyte and also secrete cytokines (e.g., tumor necrosis factor and interleukin 1) that modulate the immune process in T and B lymphocytes. Cytokines facilitate the generation of an inflammatory response, activate alveolar macrophages, and recruit additional phagocytes and other immunologic factors from plasma. The inflammatory exudate is responsible for many of the local signs of pulmonary consolidation and for the systemic manifestations of pneumonia, such as fever, chills, myalgias, and malaise.

TRANSMISSION

Microbial pathogens may enter the lung by one of several routes.

Aspiration of Organisms That Colonize the Oropharynx Most pulmonary pathogens originate in the oropharyngeal flora. Aspiration of these pathogens is the most common mechanism for the production of pneumonia. At various times during the year, healthy individuals transiently carry common pulmonary pathogens in the nasopharynx; these pathogens include *Streptococcus pneumoniae*, *S. pyogenes*, *Mycoplasma pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*. The sources of anaerobic pulmonary pathogens, such as *Porphyromonas gingivalis*, *Prevotella melaninogenica*, *Fusobacterium nucleatum*, *Actinomyces* spp., spirochetes, and anaerobic streptococci, are the gingival crevice and dental plaque, which contain more than 10^{11}colony-forming units (CFU) of microorganisms per gram. The frequency of aerobic gram-negative bacillary colonization of the oropharyngeal mucosa, which is unusual in healthy persons (<2%), increases with hospitalization, worsening debility, severe underlying illness, alcoholism, diabetes, and advanced age. This change may be a consequence of increased salivary proteolytic activity, which destroys fibronectin, a glycoprotein coating the surface of the mucosa. Fibronectin is the receptor for the normal gram-positive flora of the oropharynx. Loss of fibronectin exposes the receptors for aerobic gram-negative bacilli on the epithelial cell surface. The source of aerobic gram-negative bacilli may be the patient's own stomach (which can become colonized with these organisms as the

result of an increase in gastric pH with atrophic gastritis or after the use of H_2-blocking agents or antacids), contaminated respiratory equipment, hands of health care workers, or contaminated food and water. Nasogastric tubes can facilitate the transfer of gastric bacteria to the pharynx.

About 50% of healthy adults aspirate oropharyngeal secretions into the lower respiratory tract during sleep. Aspiration occurs more frequently and may be more pronounced in individuals with an impaired level of consciousness (e.g., alcoholics; drug abusers; and patients who have had seizures, strokes, or general anesthesia), neurologic dysfunction of the oropharynx, and swallowing disorders or mechanical impediments (e.g., nasogastric or endotracheal tubes). Pneumonia due to anaerobes is an especially likely outcome if the aspirated material is large in volume or contains virulent components of the anaerobic microbial flora or foreign bodies, such as aspirated food or necrotic tissue. Impairment of the cough reflex increases the risk of pneumonia, as does mucociliary or alveolar macrophage dysfunction.

Inhalation of Infectious Aerosols Deposition of inhaled particles within the respiratory tract is determined primarily by particle size. Particles>10 um in diameter are deposited mostly in the nose and upper airways. Particles <5 um in diameter (also called *airborne droplet nuclei*) and containing one or perhaps two microorganisms fail to settle out by gravity but rather remain suspended in the atmosphere for long periods unless removed by ventilation or by filtration in the lungs of the individual breathing the contaminated air. Transmission of an infectious agent in the form of an aerosol is particularly efficient. These infectious aerosols are small enough to bypass host defenses in the upper respiratory tract and airways. A greater percentage of particles are deposited in small bronchioles and alveoli as particle size decreases below 5 um. One inhaled particle of appropriate size may be sufficient to reach the alveolus and initiate infection. The etiologies of pneumonia typically acquired by inhalation of infectious aerosols include tuberculosis, influenza, legionellosis, psittacosis, histoplasmosis, Q fever, and hantavirus pulmonary syndrome (HPS).

Hematogenous Dissemination from an Extrapulmonary Site Infection, usually with *Staphylococcus aureus*, disseminates hematogenously to the lungs in patients (such as intravenous drug users) who have either right- or left-sided bacterial endocarditis and in patients with intravenous catheter infections. *Fusobacterium* infections of the retropharyngeal tissues (Lemierre's syndrome -- i.e., retropharyngeal abscess and jugular venous thrombophlebitis) also disseminate hematogenously to the lungs.

Direct Inoculation and Contiguous Spread Two additional routes of transmission of bacteria to the lungs are direct inoculation (as a result of either tracheal intubation or stab wounds to the chest) and contiguous spread from an adjacent site of infection.

PATHOLOGY

The pneumonic process may involve primarily the interstitium or the alveoli. Involvement of an entire lobe is called *lobar pneumonia*. When the process is restricted to alveoli contiguous to bronchi, it is called *bronchopneumonia*. Confluent bronchopneumonia may be indistinguishable from lobar pneumonia. Cavities develop when necrotic lung tissue is discharged into communicating airways, resulting in either necrotizing

pneumonia (multiple small cavities, each <2 cm in diameter, in one or more bronchopulmonary segments or lobes) or lung abscess (one or more cavities>2 cm in diameter). The classification of pneumonia is best based upon the causative microorganism rather than upon these anatomic characteristics (the criteria used in the past).

EPIDEMIOLOGY

The patient's living circumstances, occupation, travel history, pet or animal exposure history, and contacts with other ill individuals as well as the physician's knowledge of the epidemic curve of community outbreaks provide clues to the microbial etiology of a given case of pneumonia (Table 255-1). The relative frequency of various pulmonary pathogens varies with the setting in which the infection was acquired -- e.g., community, nursing home, or hospital. In patients hospitalized with community-acquired pneumonia, the most frequent pathogens are *S. pneumoniae*, *H. influenzae*, *Chlamydia pneumoniae*, and *Legionella pneumophila*. *C. pneumoniae* is often found in association with other pathogens, including *S. pneumoniae*, and the associated pathogen appears to influence the course of the pneumonia. *M. pneumoniae*, which usually causes mild illness, is common among outpatients with community-acquired pneumonia, but may also be an underappreciated cause in all age groups of severe pneumonia that requires hospitalization. In contrast, enteric aerobic gram-negative bacilli and *Pseudomonas aeruginosa*, uncommon causes of community-acquired pneumonia, are estimated to account for>50% of cases of hospital-acquired pneumonia, while *S. aureus* is responsible for >10%. The relative frequencies of pathogens in pneumonia acquired in nursing homes fall somewhere between those of community- and hospital-acquired pneumonia. Enteric aerobic gram-negative bacilli and *P. aeruginosa* are more common among nursing home residents than among patients who acquire pneumonia in noninstitutional settings.

The season of the year and the geographic location are other predictors of etiology. The frequency of influenza virus as a cause of both community-acquired and institutionally acquired pneumonia increases during the winter months. Moreover, influenza virus infection causes an increase in the frequency of secondary bacterial pneumonia due to *S. pneumoniae*, *S. aureus*, and *H. influenzae*. Outbreaks of influenza in a community tend to be explosive and widespread, with many secondary cases resulting from the short incubation period of several days and the high degree of communicability. *Legionella* colonizes hot-water storage systems that provide favorable conditions for its proliferation, such as warm temperature, stagnation, and sediment accumulation. Acquisition of *Legionella* pneumonia requires exposure to aerosols generated from these contaminated water supplies -- e.g., during an overnight stay in a hotel with a faulty air-handling system or after repair of domestic plumbing in buildings with contaminated water supplies. Legionellosis also occurs in explosive outbreaks when large numbers of susceptible people are exposed to an infectious aerosol; however, no secondary cases occur because of the low level of communicability of *L. pneumophila*. *Mycoplasma* causes outbreaks, usually in relatively closed populations such as those at military bases, at colleges, or in households; however, because of its long incubation period (2 to 3 weeks) and its relatively low degree of communicability, *Mycoplasma* infection moves through the community slowly, affecting another person as the first is recovering. In communities where infection with HIV type 1 is endemic, *Pneumocystis*

carinii and *M. tuberculosis* are more prominent causes of community-acquired pneumonia. *Chlamydia psittaci* produces illness in bird handlers. Histoplasmosis, blastomycosis, and coccidioidomycosis are causes of pneumonia that have specific geographic distributions.

HPSis a newly described, frequently fatal disease caused by one of several hantaviruses. Most cases in the United States have been reported from the Four Corners area (New Mexico, Arizona, Utah, and Colorado), where the pathogen is the Sin Nombre virus. The primary hosts are rodents, which apparently remain healthy but excrete the virus in urine, feces, and saliva. Hantavirus infection is acquired by inhalation of infectious aerosols when rodent nests are disturbed by human domestic, occupational, or recreational activities. The appearance of HPS in the southwestern United States is thought to have occurred because of increased rainfall in the region, which increased the rodent food supply and thus the rodent population. No person-to-person transmission of HPS is thought to have taken place, except perhaps in an outbreak in southern Argentina in 1996.

AGE AND COMORBIDITY

Age is an important predictor of the infecting agent in pneumonia. *Chlamydia trachomatis* and respiratory syncytial virus are common among infants< 6 months of age; *H. influenzae* among children 6 months to 5 years of age; *M. pneumoniae*, *C. pneumoniae,* and hantavirus among young adults; *H. influenzae* and *M. catarrhalis* among elderly individuals with chronic lung disease; and *L. pneumophila* among elderly persons, smokers, and persons with compromised cell-mediated immunity (e.g., transplant recipients), renal or hepatic failure, diabetes, or systemic malignancy.

Oral anaerobes, frequently in combination with aerobic bacterial components of the human flora (e.g., viridans streptococci), are causes of community-acquired pneumonia and anaerobic lung abscess in patients who are prone to aspiration. Edentulous persons, who have lower numbers of oral anaerobes, are less likely to develop pneumonia due to anaerobes. When the etiology of community-acquired pneumonia in unselected hospitalized patients has been studied by methods that entail strict anaerobic bacteriology and that avoid contamination of lower respiratory tract secretions by the oral flora, anaerobic bacteria have been found to account for as many as 20 to 30% of cases. In hospital-acquired pneumonia, anaerobes are the pathogens -- with or without aerobic copathogens -- in about one-third of cases. However, the aerobic copathogens in hospital-acquired pneumonia are frequently virulent microorganisms in their own right (e.g., enteric aerobic gram-negative bacilli, *P. aeruginosa*, and *S. aureus*).

The patient's underlying disease may be characterized by specific immunologic or inflammatory defects that predispose to pneumonia due to specific pathogens (Table 255-2). For example, immunoglobulin deficiencies -- especially those involving IgG subtypes 2 and 4, which are important in the immune response to encapsulated organisms (e.g., *S. pneumoniae* and *H. influenzae*) -- may be associated with recurrent sinopulmonary infections. Immunoglobulin deficiencies may be inherited, or they may be acquired (i.e., as a result of either decreased production, as in lymphoproliferative malignancies, or excessive protein loss, as in nephrosis or protein-losing enteropathy).

Inherited immunoglobulin deficiencies may be global or selective. Patients with recurrent sinopulmonary infections and a selective deficiency of IgG2 and/or IgG4 may have a total plasma IgG level within the normal range, as these particular IgG subtypes constitute only 25% of total IgG. HIV-infected patients may also exhibit ineffective antibody formation, which predisposes to infection with these encapsulated bacteria. Severe neutropenia (<500 neutrophils/uL) increases the risk of infections due to *P. aeruginosa*, Enterobacteriaceae, *S. aureus*, and (if neutropenia is prolonged) *Aspergillus*. The risk is unusually high for infections due to *M. tuberculosis* among HIV-infected patients with circulating CD4+ lymphocyte counts of<500/uL; for infections due to *P. carinii*, *Histoplasma capsulatum*, and *Cryptococcus neoformans* among those with CD4+ counts of<200/uL; and for infections due to *M. avium-intracellulare* and cytomegalovirus among those with counts of<50/uL. Long-term glucocorticoid therapy increases the risk of tuberculosis and nocardiosis.

CLINICAL MANIFESTATIONS

Community-Acquired Pneumonia Community-acquired pneumonia has traditionally been thought to present as either of two syndromes: the typical presentation or the atypical presentation. Although current data suggest that these two syndromes may be less distinct than was once thought, the characteristics of the clinical presentation may nevertheless have some diagnostic value.

The "typical" pneumonia syndrome is characterized by the sudden onset of fever, cough productive of purulent sputum, shortness of breath, and (in some cases) pleuritic chest pain; signs of pulmonary consolidation (dullness, increased fremitus, egophony, bronchial breath sounds, and rales) may be found on physical examination in areas of radiographic abnormality. The typical pneumonia syndrome is usually caused by the most common bacterial pathogen in community-acquired pneumonia, *S. pneumoniae*, but can also be due to other bacterial pathogens, such as *H. influenzae* and mixed anaerobic and aerobic components of the oral flora.

The "atypical" pneumonia syndrome is characterized by a more gradual onset, a dry cough, shortness of breath, a prominence of extrapulmonary symptoms (such as headache, myalgias, fatigue, sore throat, nausea, vomiting, and diarrhea), and abnormalities on chest radiographs despite minimal signs of pulmonary involvement (other than rales) on physical examination. Atypical pneumonia is classically produced by *M. pneumoniae* but can also be caused by *L. pneumophila*, *C. pneumoniae*, oral anaerobes, and *P. carinii* as well as by *S. pneumoniae* and the less frequently encountered pathogens *C. psittaci*, *Coxiella burnetii*, *Francisella tularensis*, *H. capsulatum*, and *Coccidioides immitis*. *Mycoplasma* pneumonia (Chap. 178) may be complicated by erythema multiforme, hemolytic anemia, bullous myringitis, encephalitis, and transverse myelitis. *Legionella* pneumonia (Chap. 151) is frequently associated with deterioration in mental status, renal and hepatic abnormalities, and marked hyponatremia; pneumonia due to *H. capsulatum* (Chap. 201) or *C. immitis* (Chap. 202) is often accompanied by erythema nodosum. In *C. pneumoniae* pneumonia (Chap. 179), sore throat, hoarseness, and wheezing are relatively common. The atypical pneumonia syndrome in patients whose behavioral history places them at risk of HIV infection suggests *Pneumocystis* infection. These patients may have concurrent infections caused by other opportunistic pathogens, such as pulmonary (and frequently

extrapulmonary) tuberculosis, oral thrush due to *Candida albicans*, or extensive perineal ulcers due to herpes simplex virus.

Certain viruses also produce pneumonia that is usually characterized by an atypical presentation -- i.e., chills, fever, shortness of breath, dry nonproductive cough, and predominance of extrapulmonary symptoms. Primary viral pneumonia can be caused by influenza virus (usually as part of a community outbreak in winter), by respiratory syncytial virus (in children and immunosuppressed individuals), by measles or varicella-zoster virus (accompanied by the characteristic rash), and by cytomegalovirus (in patients immunocompromised by HIV infection or by therapy given in association with organ transplantation). Hantavirus causes an initial nonspecific febrile prodrome, after which the patient develops rapidly progressive respiratory failure and diffuse pulmonary infiltrates on chest radiographs as a result of exudation into the pulmonary interstitium and alveoli, with thrombocytopenia, neutrophilic leukocytosis, circulating immunoblasts, and laboratory evidence of hemoconcentration. In addition, influenza and measles can predispose to secondary bacterial pneumonia as a result of the destruction of the mucociliary barrier of the airways. Secondary bacterial infection may either follow the viral infection without interruption or be separated from the viral infection by several days of transient relief of symptoms. Bacterial infection may be heralded by sudden worsening of the patient's clinical condition, with persisting or renewed chills, fever, and cough productive of purulent sputum, possibly accompanied by pleuritic chest pain.

Patients with hematogenous *S. aureus* pneumonia may present with fever and dyspnea only. In these cases the inflammatory response is initially confined to the pulmonary interstitium. Cough, sputum production, and signs of pulmonary consolidation develop only after the infection extends into the bronchi. These patients are usually gravely ill, with intravascular infection as well as pneumonia, and may have signs of endocarditis (Chap. 126).

Nocardiosis (Chap. 165) is frequently complicated by metastasis of lesions to the skin and central nervous system. Signs of pulmonary consolidation, cough, and sputum production may be lacking in patients who are unable to mount an inflammatory response, such as those with agranulocytosis. The major manifestations in these patients may be limited to fever, tachypnea, agitation, and altered mental status. Elderly or severely ill patients may fail to develop fever.

Tuberculosis also produces an atypical presentation that is characterized by fever, night sweats, cough, and shortness of breath and sometimes by pleuritic chest pain and blood-streaked sputum. Several weeks usually elapse before the patient seeks medical attention because of the gradual worsening of these symptoms, by which time he or she will have lost considerable weight.

Nosocomial Pneumonia Patients with nosocomial pneumonia often pose a diagnostic challenge. The differential diagnosis of acute respiratory disease in critically ill, hospitalized patients is diverse and includes noninfectious entities, such as congestive heart failure, acute respiratory distress syndrome, preexisting lung disease, atelectasis, and oxygen- or drug-related toxicities, that may be difficult to distinguish clinically or radiologically from pneumonia. The usual criteria for nosocomial pneumonia, which include new or progressive pulmonary infiltrates, purulent tracheobronchial secretions,

fever, and leukocytosis, are frequently unreliable in these patients, who often have preexisting pulmonary disease, endotracheal tubes that irritate the tracheal mucosa and may elicit an inflammatory exudate in respiratory secretions, or multiple other problems likely to produce fever and leukocytosis. Patients with nosocomial pneumonia complicating an underlying illness associated with significant neutropenia often have no purulent respiratory tract secretions or pulmonary infiltrates, and patients with nosocomial pneumonia complicating uremia or cirrhosis often remain afebrile. In addition, the patients at greatest risk for nosocomial pneumonia are most likely to be heavily colonized with potential pulmonary pathogens in the oropharyngeal or tracheobronchial mucosa; thus the presence of these organisms in gram-stained preparations or cultures of respiratory tract secretions does not necessarily confirm the diagnosis of pneumonia.

Aspiration Pneumonia and Anaerobic Lung Abscess Aspiration of a sufficient volume of gastric acid produces a chemical pneumonitis characterized by acute dyspnea and wheezing with hypoxemia and infiltrates on chest radiographs in one or both lower lobes. Clinical findings following aspiration of particulate matter depend on the extent of endobronchial obstruction and range from acute apnea to persistent cough with or without recurrent infection. Although the aspiration of oral anaerobes can initially lead to an infiltrative process, it ultimately results in putrid sputum, tissue necrosis, and pulmonary cavities. In about three-quarters of cases, the clinical course of an abscess of anaerobic polymicrobial etiology is indolent and mimics that of pulmonary tuberculosis, with cough, shortness of breath, chills, fever, night sweats, weight loss, pleuritic chest pain, and blood-streaked sputum lasting for several weeks or more. In other patients the disease may present more acutely. Patients with anaerobic abscesses are usually prone to aspiration of oropharyngeal contents and have periodontal disease. One genus of oral anaerobes, *Actinomyces*, produces a chronic fibrotic necrotizing process that crosses tissue planes and may involve the pleural space, ribs, vertebrae, and subcutaneous tissue, with eventual discharge of sulfur granules (macroscopic bacterial masses) through the skin (empyema necessitatis).

DIAGNOSIS

Radiography Chest radiography is more sensitive than physical examination for detection of pulmonary infiltrates. Indeed, *P. carinii* pneumonia (PCP) is the only relatively common form of pneumonia associated with false-negative chest radiographs; up to 30% of patients with PCP have false-negative results. Chest radiographs can confirm the presence and location of the pulmonary infiltrate; assess the extent of the pulmonary infection; detect pleural involvement, pulmonary cavitation, or hilar lymphadenopathy; and gauge the response to antimicrobial therapy. However, chest radiographs may be normal when the patient is unable to mount an inflammatory response (e.g., in agranulocytosis) or is in the early stage of an infiltrative process (e.g., in hematogenous *S. aureus* pneumonia or PCP associated with AIDS). High-resolution computed tomography of the lungs can improve the accuracy of diagnosis of pneumonia, especially when the process involves lung obscured by the diaphragm, liver, ribs and clavicles, or heart.

The anatomic localization of the inflammatory process, as visualized in chest radiographs, occasionally has diagnostic implications. Most pulmonary pathogens

produce focal lesions. A multicentric distribution suggests hematogenous infection, in which case the remote location of the primary infection (e.g., endocarditis or thrombophlebitis) should be sought. Hematogenous pneumonia, which results from septic embolization in patients with thrombophlebitis or right-sided endocarditis and from bacteremia in patients with left-sided endocarditis, appears on the chest radiograph as multiple areas of pulmonary infiltration that subsequently may cavitate. A diffuse distribution suggests the involvement of *P. carinii*, cytomegalovirus, hantavirus, measles virus, or herpes zoster virus (with pneumonia due to the last two pathogens diagnosed by the characteristic accompanying rash). Pleurisy and hilar nodal enlargement are unusual with PCPand cytomegalovirus pneumonia; their presence suggests another etiology. Diffuse lesions in immunocompromised patients also suggest legionellosis, tuberculosis, histoplasmosis, *Mycoplasma* infection, or disseminated strongyloidiasis.

Oral anaerobes, *S. aureus, S. pneumoniae* serotype III, aerobic gram-negative bacilli, *M. tuberculosis*, and fungi as well as certain noninfectious conditions can produce tissue necrosis and pulmonary cavities (Table 255-3). In contrast, *H. influenzae*, *M. pneumoniae*, viruses, and most other serotypes of *S. pneumoniae* almost never cause cavities. Apical disease, with or without cavities, suggests reactivation tuberculosis. Anaerobic abscesses are located in dependent, poorly ventilated, and poorly draining bronchopulmonary segments and characteristically have air-fluid levels, unlike the well-ventilated, well-drained upper-lobe cavities caused by *M. tuberculosis*, an obligate aerobe. Air-fluid levels may also be present in cavities due to pulmonary necrosis of other infectious etiologies, such as *S. aureus* and aerobic gram-negative bacilli. *Mucor* and *Aspergillus* invade blood vessels and cause pleural-based, wedge-shaped areas of pulmonary infarction; these infarcts may subsequently cavitate.

In the patient with an uncomplicated course, chest radiographs need not be repeated before discharge, since the resolution of infiltrates may take up to 6 weeks after initial presentation. However, patients who do not respond clinically, who have a pleural effusion on admission, who may have postobstructive pneumonia, or who are infected with certain pathogens (e.g., *S. aureus*, aerobic gram-negative bacilli, or oral anaerobes) need more intensive surveillance. At times, computed tomography may be especially helpful in distinguishing different processes -- e.g., pleural effusion versus underlying pulmonary consolidation, hilar adenopathy versus pulmonary mass, and pulmonary abscess versus empyema with an air-fluid level.

Sputum Examination Examination of the sputum remains the mainstay of the evaluation of a patient with acute bacterial pneumonia. Unfortunately, expectorated material is frequently contaminated by potentially pathogenic bacteria that colonize the upper respiratory tract (and sometimes the lower respiratory tract) without actually causing disease. This contamination reduces the diagnostic specificity of any lower respiratory tract specimen. In addition, it has been estimated that the usual laboratory processing methods detect the pulmonary pathogen in fewer than 50% of expectorated sputum samples from patients with bacteremic *S. pneumoniae* pneumonia. This low sensitivity may be due to misidentification of the a-hemolytic colonies of *S. pneumoniae* as nonpathogenic a-hemolytic streptococci ("normal flora"), overgrowth of the cultures by hardier colonizing organisms, or loss of more fastidious organisms due to slow transport or improper processing. In addition, certain common pulmonary pathogens, such as anaerobes, mycoplasmas, chlamydiae, *Pneumocystis*, mycobacteria, fungi, and

legionellae, cannot be cultured by routine methods.

Since expectorated material is routinely contaminated by oral anaerobes, the diagnosis of anaerobic pulmonary infection is frequently inferred. Confirmation of such a diagnosis requires the culture of anaerobes from pulmonary secretions that are uncontaminated by oropharyngeal secretions, which in turn requires the collection of pulmonary secretions by special techniques, such as transtracheal aspiration (TTA), transthoracic lung puncture, and protected brush via bronchoscopy. These procedures are invasive and are usually not used unless the patient fails to respond to empirical therapy.

Gram's staining of sputum specimens, screened initially under low-power magnification (10´ objective and 10´ eyepiece) to determine the degree of contamination with squamous epithelial cells, is of utmost diagnostic importance. In patients with the typical pneumonia syndrome who produce purulent sputum, the sensitivity and specificity of Gram's staining of sputum minimally contaminated by upper respiratory tract secretions (>25 polymorphonuclear leukocytes and<10 epithelial cells per low-power field) in identifying the pathogen as *S. pneumoniae* are 62 and 85%, respectively. Gram's staining in this case is more specific and probably more sensitive than the accompanying sputum culture. The finding of mixed flora on Gram's staining of an uncontaminated sputum specimen suggests an anaerobic infection. Acid-fast staining of sputum should be undertaken when mycobacterial infection is suspected. Examination by an experienced pathologist of Giemsa-stained expectorated respiratory secretions from patients with AIDS has given satisfactory results in the diagnosis ofPCP. The sensitivity of sputum examination is enhanced by the use of monoclonal antibodies to *Pneumocystis* and is diminished by prior prophylactic use of inhaled pentamidine. Blastomycosis can be diagnosed by the examination of wet preparations of sputum. Sputum stained directly with fluorescent antibody can be examined for *Legionella*, but this test yields false-negative results relatively often. Thus sputum should also be cultured for *Legionella* on special media.

Expectorated sputum usually is easily collected from patients with a vigorous cough but may be scant in patients with an atypical syndrome, in the elderly, and in persons with altered mental status. If the patient is not producing sputum and can cooperate, respiratory secretions should be induced with ultrasonic nebulization of 3% saline. An attempt to obtain lower respiratory secretions by passage of a catheter through the nose or mouth rarely achieves the desired results in an alert patient and is discouraged; usually the catheter can be found coiled in the oropharynx.

In some cases that do not require the patient's hospitalization (see "Decision to Hospitalize," below), an accurate microbial diagnosis may not be crucial, and empirical therapy can be started on the basis of clinical and epidemiologic evidence alone. This approach may also be appropriate for hospitalized patients who are not severely ill and who are unable to produce an induced sputum specimen. Use of invasive procedures to establish a microbial diagnosis carries risks that must be weighed against potential benefits. However, the decision to initiate empirical therapy without an evaluation of induced sputum should be undertaken with caution and, in the case of hospitalized patients, should always be accompanied by the culture of several blood samples. The ability to understand the cause of a poor response to empirical antimicrobial therapy (Table 255-4) may be compromised by the lack of initial sputum and blood cultures.

Establishing a specific microbial etiology in the individual patient is important, for it allows institution of specific pathogen-directed antimicrobial therapy and reduces the use of broad-spectrum combination regimens to cover multiple possible pathogens. Use of a single narrow-spectrum antimicrobial agent exposes the patient to fewer potential adverse drug reactions and reduces the pressure for selection of antimicrobial resistance. Emergence of antimicrobial resistance is a type of adverse drug reaction unlike others, because it is "contagious." In addition, establishing a microbial diagnosis can help define local community outbreaks and antimicrobial resistance patterns.

Invasive Procedures The sensitivities and specificities of the invasive procedures described below for obtaining pulmonary material vary with the type of immunocompromised patient, the type of pulmonary lesion, and the degree of prior exposure to therapeutic or prophylactic antimicrobial agents.

Transtracheal Aspiration Popular several decades ago,TTA is rarely performed today. Although the sensitivity of the procedure is high (approaching 90%), the specificity is low. The material obtained by TTA (from a catheter inserted through the cricothyroid cartilage and advanced toward the carina) is not contaminated by upper respiratory tract secretions but can contain organisms that colonize the tracheobronchial tree without necessarily causing pneumonia. Significant morbidity and even death have attended the use of TTA. Contraindicated in patients with a bleeding diathesis, TTA may cause infection at the puncture site and may lead to severe subcutaneous and mediastinal emphysema in patients who are coughing vigorously.

Percutaneous Transthoracic Lung Puncture This procedure employs a skinny (small-gauge) needle that is advanced into the area of pulmonary consolidation with computed tomographic guidance. It requires that the patient cooperate, have good hemostasis, and be able to tolerate a possible associated pulmonary hemorrhage or pneumothorax. Patients on mechanical ventilation cannot undergo lung puncture because of the high incidence of complicating pneumothorax.

Fiberoptic Bronchoscopy Fiberoptic bronchoscopy is safe and relatively well tolerated and has become the standard invasive procedure used to obtain lower respiratory tract secretions from seriously ill or immunocompromised patients with complex or progressive pneumonia. This technique provides a direct view of the lower airways. Specimens obtained by bronchoscopy should be subjected to Gram's, acid-fast, *Legionella*direct fluorescent antibody, and Gomori's methenamine silver staining and should be cultured for routine aerobic and anaerobic bacteria, legionellae, mycobacteria, and fungi. Samples are collected with a protected double-sheathed brush (PSB), by bronchoalveolar lavage (BAL), or by transbronchial biopsy (TBB) at the site of pulmonary consolidation. The PSB sample is usually contaminated by oropharyngeal flora; quantitative cultures of the 1 mL of sterile culture medium into which the brush is placed after withdrawal from the inner catheter must be performed to differentiate contamination (<1000CFU/mL) from infection (³1000 CFU/mL). The results of PSB are highly specific and highly sensitive, especially when the patient has not received antibiotics before culture. BAL is usually performed with 150 to 200 mL of sterile, nonbacteriostatic saline. When used to facilitate endoscopy, local anesthetic agents with antibacterial activity can lower the sensitivity of culture results. Quantitative bacteriologic evaluation of BAL fluid has given results similar to those obtained with the PSB

technique. Gram's staining of the cytocentrifuged BAL fluid specimen can serve as an immediate guide in the selection of antimicrobial therapy to be administered while culture results are awaited.

Open-Lung Biopsy This procedure is most commonly needed when specimens obtained bronchoscopically from an immunocompromised patient with progressive pneumonia have been unrevealing. Limitations on the performance of an open-lung biopsy include hypoxemia and a bleeding diathesis, which may supervene while the physician is deciding whether to undertake this procedure. Results of an open-lung biopsy are considered diagnostic because of the large size of the tissue sample. The diagnostic yield of this procedure is greatest in focal lesions, whereas bronchoscopic evaluation is most useful in diffuse lesions.

Other Diagnostic Tests In the initial evaluation of a patient with pneumonia, at least two blood samples for culture should be obtained from different venipuncture sites; if empyema is a clinical consideration, diagnostic thoracentesis is indicated. Positive blood or pleural fluid culture is generally considered diagnostic of the etiology of pneumonia. However, bacteremia and empyema each occur in fewer than 10 to 30% of patients with pneumonia.

Serologic studies are sometimes helpful in defining the etiology of certain types of pneumonia, although serologic diagnosis -- because it is often delayed by the need to demonstrate at least a fourfold rise in convalescent-phase antibody titer -- is usually retrospective. A single IgM antibody titer of >1:16, a single IgG antibody titer of >1:128, or a fourfold or greater rise in the IgG titer obtained by indirect immunofluorescence is diagnostic of *M. pneumoniae* infection. A single IgM antibody titer of³1:20, a single IgG antibody titer of³1:128, or a fourfold or greater rise in the IgG titer obtained by micro-indirect immunofluorescence is diagnostic of *C. pneumoniae* infection. A single *Legionella* antibody titer of³1:256 or a fourfold rise to a titer of³1:128 suggests acute legionellosis. A highly sensitive and specific urinary antigen test is available to detect *L. pneumophila* serogroup 1 in patients with pneumonia; this organism accounts for ~70% of *L. pneumophila* infections. The diagnosis of hantavirus infection is confirmed by detection of IgM serum antibodies, a rising titer of IgG serum antibodies, hantavirus-specific RNA by polymerase chain reaction in clinical specimens, and hantavirus-specific antigen by immunohistochemistry.

DECISION TO HOSPITALIZE

Approximately 20% of patients with community-acquired pneumonia are hospitalized, some perhaps unnecessarily. Use of inpatient hospital services is costly and at times poses risks to the patient (e.g., the risk of nosocomial infections). Thus hospitalization must be justified by anticipation of a poor outcome if the case is managed in an outpatient setting.

The Pneumonia Patient Outcomes Research Team (PORT) has attempted to quantify the risk of death and other adverse outcomes of community-acquired pneumonia by assignment of points to 19 variables (Fig. 255-1), with stratification of patients into five classes based on cumulative point score. This prediction rule was derived and validated in a large number of patients. On the basis of their observations, the PORT investigators

suggest that outpatient management is appropriate for many patients in classes I and II, in whom the risks of subsequent hospitalization (£8.2%) and of death (<0.6%) are low. They suggest outpatient management after a short hospital stay for patients in class III, whose risk of subsequent hospitalization if initially treated at home is 16.7% but whose risk of admission to the intensive care unit (ICU) is 5.9% -- similar to that for patients in classes I and II. The PORT investigators further suggest that patients in classes IV and V (risk of death, 8.2 and 29.2%, respectively; risk of ICU admission, 11.4 and 17.3%, respectively) should receive traditional inpatient care. An expert panel from the Infectious Diseases Society of America (IDSA) endorses the PORT recommendations.

Other characteristics that favor a decision to hospitalize the patient include the known presence of certain etiologic microorganisms (e.g., *S. aureus*) that are associated with a poor prognosis, multilobe pulmonary involvement, suppurative complications (e.g., empyema or septic arthritis), evidence of poor functional status (e.g., hypotension or hypoxemia on presentation in patients otherwise in classes I, II, and III), evidence of a patient's inability to comply with treatment recommendations, anticipated difficulty in assessing the response to outpatient treatment, and an inadequate home support system that may compromise outpatient care. Discharge from the hospital should be guided by similar considerations.

TREATMENT

Community-Acquired Pneumonia: Outpatient Management Most cases of community-acquired pneumonia in otherwise-healthy adults do not require hospitalization. Although desirable, it is often impractical in the outpatient setting to obtain a chest radiograph and sputum Gram's stain and culture in order to confirm the clinical diagnosis of pneumonia and its microbial etiology before starting antimicrobial therapy. Consequently, the oral antimicrobial treatment administered in the outpatient setting is frequently empirical (Table 255-5). The pathogen in such a situation is likely to be *M. pneumoniae*, *S. pneumoniae*, or *C. pneumoniae*. In older patients with underlying chronic respiratory disease, *L. pneumophila*, *H. influenzae*, or *M. catarrhalis* should also be considered. In patients at risk of aspiration, oral anaerobes may be involved. Few oral antimicrobial drugs have a reliable spectrum encompassing all of these pathogens (Table 255-5). Whatever regimen is chosen, its antimicrobial activity should encompass *S. pneumoniae*, the most common cause of pneumonia. Increasing resistance among pneumococci to all the available oral antimicrobial agents precludes the designation of any one agent as the clear drug of choice.

Strains of *S. pneumoniae* for which the minimal inhibitory concentration (MIC) of penicillin (as determined by the broth dilution method) is 0.1 to 1.0 ug/mL are considered to have intermediate-level resistance, while strains whose MIC is>1.0 ug/mL are considered to have high-level resistance. The current, less time-consuming method to screen for penicillin resistance is the use of a 1-ug oxacillin disk in a disk diffusion assay. Penicillin resistance (i.e., an MIC ³0.1 ug/mL) is indicated by a zone of growth inhibition of£19 mm. Antimicrobial gradient paper strips (the E-test), which yield the exact MIC, are as accurate as the broth dilution technique, can be performed as rapidly as the oxacillin disk diffusion assay, and have replaced the oxacillin disk test in many institutions.

The resistance of *S. pneumoniae* to penicillin varies greatly with the source of the clinical sample tested (e.g., strains isolated from middle-ear fluid are most often resistant), the age of the patient (e.g., resistance is more frequent among children than among adults), the setting (e.g., resistance is more common in day-care centers), the patient's socioeconomic status (the frequency of resistance is highest in samples from suburban and white patients), and the geographic region in which the specimen was collected. Caution must be exercised in the interpretation of surveys of antimicrobial resistance among pneumococci in the United States, which can be strongly affected by these types of sampling bias. In a national survey of clinical isolates from normally sterile body sites that was conducted in 1997 in various surveillance areas throughout the United States by the Centers for Disease Control and Prevention (CDC), 11% (range, 6 to 19%) of 3110 isolates of *S. pneumoniae* exhibited intermediate-level resistance to penicillin, and 14% (range, 8 to 26%) displayed high-level resistance. However, in another national survey of the antimicrobial susceptibility of clinical isolates obtained from respiratory tract sites between February and June 1997 at 27 U.S. medical centers (SENTRY surveillance program), 28% of 845 isolates (with a range of 11 to 52% at the various medical centers) displayed intermediate-level penicillin resistance, and an additional 16% (with a range of 0 to 33%) displayed high-level penicillin resistance.

As a consequence of the production of altered penicillin-binding proteins with decreasedb-lactam affinity, penicillin-resistant *S. pneumoniae* exhibits at least some degree of cross-resistance to allb-lactams, including the extended-spectrum third- and fourth-generation cephalosporins. Since the mechanism of penicillin resistance does not involve b-lactamase production,b-lactam/b-lactamase inhibitor combinations (e.g., amoxicillin/clavulanate) offer no advantage. Indeed, theMICs of penicillin and amoxicillin are nearly identical, but the serum levels after equivalent doses are much higher for amoxicillin than for penicillin, a difference that may reflect a therapeutic advantage of amoxicillin. Among the oral cephalosporins, cefaclor, cefadroxil, and cephalexin have variable activity against penicillin-sensitive strains; cefuroxime and cefpodoxime have activity against penicillin-susceptible strains but variable activity against penicillin-intermediate strains and no activity against highly penicillin-resistant strains.

Resistance to other antimicrobial agents, such as the macrolides (erythromycin, clarithromycin, and azithromycin), clindamycin, tetracycline and doxycycline, and trimethoprim-sulfamethoxazole (TMP-SMZ), is also more common among penicillin-intermediate strains than among penicillin-susceptible strains, and it is most common among highly penicillin-resistant strains. Overall rates of resistance among *S. pneumoniae* strains are ~14% for the macrolides, 4% for clindamycin, up to 10% for tetracyclines, and 20 to 30% for TMP-SMZ. Rates of resistance to the newer fluoroquinolones levofloxacin, gatifloxacin, moxifloxacin, and sparfloxacin are <4%, regardless of penicillin susceptibility. At best, the older fluoroquinolones (e.g., ciprofloxacin) have borderline activity, as judged by serum levels in relation to MICsof these drugs against the pneumococcus.

Optimally, the choice of antimicrobial drugs for empirical therapy should be guided by local resistance patterns, if known. Options for empirical antimicrobial therapy should be modified in light of continually evolving antimicrobial resistance patterns resulting from the introduction of new resistant clones into the community from other regions or the

emergence of resistant mutants under the selective pressure of local patterns of antimicrobial use. TheIDSA has published guidelines for the treatment of community-acquired pneumonia. These guidelines emphasize the need for a chest radiograph when pneumonia is suspected and for the establishment of a microbial diagnosis (e.g., by sputum Gram's stain with or without culture) whenever possible. Doxycycline and the newer fluoroquinolones are recommended alternatives for initial empirical oral therapy, especially when penicillin-resistant pneumococci are suspected. The utility of the macrolides and amoxicillin depends on susceptibility of pneumococci in the local community.

The regimen should be modified for patients with particular epidemiologic factors or comorbidities related to specific pathogens\em\e.g., structural lung disease or suspected aspiration. Aspiration pneumonia can be treated with amoxicillin/clavulanate, clindamycin, or amoxicillin plus metronidazole because these regimens are active against oral anaerobes. Metronidazole alone has inadequate activity against microaerophilic gram-positive cocci and must be supplemented with a b-lactam agent that compensates for this defect in spectrum. If macrolides are used and *H. influenzae* is suspected, azithromycin or clarithromycin is preferred because of erythromycin's poor activity against this organism. Alternative agents for *H. influenzae* include amoxicillin/clavulanate, doxycycline, or a fluoroquinolone. Theb-lactams are not active against pathogens causing atypical pneumonia (e.g., *Mycoplasma*, *C. pneumoniae*, or *Legionella*), in which case doxycycline, a macrolide, or a fluoroquinolone is preferred.

TheIDSAguidelines recommend that pneumococcal pneumonia be treated for 7 to 10 days or until the patient has been afebrile for 72 h. Pneumonia caused by *Legionella, C. pneumoniae*, or *Mycoplasma* should be treated for 2 to 3 weeks unless azithromycin is used, in which case a 5-day course is acceptable because of the drug's prolonged half-life in tissues.

Community-Acquired Pneumonia: Inpatient Management Patients who have community-acquired pneumonia and are ill enough to be hospitalized (Fig. 255-1) must have a chest radiograph to establish the diagnosis of pneumonia, must undergo prompt microbiologic evaluation (including Gram's staining and culture of sputum and culture of two blood samples drawn by separate venipuncture), and must receive empirical antimicrobial therapy based on Gram's staining of sputum and knowledge of the current antimicrobial sensitivities of the pulmonary pathogens in the local geographic area (Tables 255-6 and255-7). Antimicrobial therapy should be initiated promptly (e.g., within 8 h of admission). Parenteral antimicrobial therapy in the hospitalized patient is usually mandatory. A lack of sputum production, an atypical clinical presentation, the presence of diffuse radiographic infiltrates, a rapidly progressive downhill course, and a poor response to prior empirical therapy are among the indications for the use of invasive procedures to detect the pulmonary pathogen, especially in the immunocompromised patient. Although broad-spectrum antibacterial therapy should be started during a full evaluation in severely ill patients with rapidly progressing illness, these empirical regimens cannot encompass all the possible pathogens without producing unnecessary toxicity and expense. Indeed, in immunocompromised patients (including those with neutropenia or HIV infection), the number of microbial and noninfectious causes of pulmonary disease is large and increasing. Since failure to provide specific treatment can prove rapidly fatal, a diagnosis should be sought aggressively so that optimal

therapy can be started promptly.

Penicillin or ampicillin remains the drug of choice for infection due to penicillin-susceptible pneumococci. Studies suggest that high-dose intravenous penicillin G (e.g., 10 to 20 million units daily), ampicillin (2 g every 6 h), ceftriaxone (1 or 2 g every 24 h), or cefotaxime (1 to 2 g every 6 h) constitutes adequate therapy for pneumonia due to strains exhibiting intermediate resistance to penicillin (MIC, 0.1 to 1 ug/mL). The effectiveness of high-dose intravenous penicillin against pneumonia due to highly resistant pneumococcal strains is unknown, but MICs of cefotaxime and ceftriaxone for these strains are usually lower than those of penicillin or ampicillin and most other b-lactam antibiotics. Ceftriaxone or cefotaxime may be effective when the MIC of penicillin is ³1 ug/mL and those of ceftriaxone and cefotaxime are £2 ug/mL. However, highly cephalosporin-resistant strains have become a problem in certain geographic areas. Since all penicillin-resistant strains are sensitive to vancomycin, initial empirical therapy should include this antibiotic (1 g intravenously every 12 h) when the patient with pneumococcal pneumonia is severely ill, has significant comorbidity, and lives in a region where highly penicillin- or cephalosporin-resistant strains have become common.

If the result of Gram's staining of sputum is not interpretable or not available, then theIDSAguidelines recommend empirical therapy for patients hospitalized on a general medical unit with a b-lactam (e.g., ceftriaxone, cefotaxime) or ab-lactam/b-lactamase inhibitor combination, with or without a macrolide, or with one of the fluoroquinolones alone. Seriously ill patients who are hospitalized in theICUshould always receive a macrolide or a newer fluoroquinolone in addition to theb-lactam to cover *Legionella*. The therapeutic regimens should be modified further in the following situations: structural disease of the lung (e.g., bronchiectasis) requires treatment with an anti-*Pseudomonas* b-lactam plus a macrolide or with a newer fluoroquinolone plus an aminoglycoside; penicillin allergy requires treatment with a newer fluoroquinolone, with or without clindamycin; and suspected aspiration requires treatment with a newer fluoroquinolone plus either clindamycin or metronidazole or with ab-lactam/b-lactamase inhibitor combination alone. A recent study of almost 13,000 elderly hospitalized patients with pneumonia, which controlled for severity of illness, baseline differences in patient characteristics, and processes of care, documented 30-day mortality that was 26 to 36% lower among those treated initially with a fluoroquinolone alone or a macrolide combined with a second- or nonpseudomonal third-generation cephalosporin than among those initially given a nonpseudomonal third-generation cephalosporin alone. This result may reflect the importance of pathogens such as *Mycoplasma*, *Legionella*, and *C. pneumoniae* in these patients.

Therapy can be switched from intravenous to oral agents within 3 days to complete a 7- to 10-day course if the patient's clinical condition improves rapidly and if antimicrobial agents that are readily absorbed after oral administration and that reach tissue levels above theMIC are available. The presence of *S. aureus* or aerobic gram-negative bacilli or the development of suppurative complications requires a more prolonged course of therapy. Pneumonia caused by *Legionella*, *C. pneumoniae*, or *Mycoplasma* should be treated for 2 to 3 weeks unless azithromycin is used. Anaerobic lung abscess should be treated with the regimens suggested for aspiration pneumonia until a chest radiograph (with radiography performed at 2-week intervals) is clear or shows only a small stable

scar. Therapy is prolonged for ³6 weeks to prevent relapse, although shorter courses are probably sufficient for many patients. Surgery is rarely required for lung abscess; indications for surgery include massive hemoptysis and suspected neoplasm. Supportive measures include the administration of supplemental oxygen and intravenous fluids, assistance in clearing secretions, fiberoptic bronchoscopy, and (if necessary) ventilatory support. Caution should be exercised in bronchoscopic drainage of large, fluid-filled lung abscesses because of the potential for sudden massive spillage of large collections of pus into the airways.

Patients with risk factors for HIV infection and an atypical pneumonia syndrome should be evaluated forPCPbecause of its frequency as an index diagnosis in HIV infection and its potential severity. Tuberculosis and other causes of atypical pneumonia must be excluded as part of the evaluation of these patients. Empirical therapy can consist of eitherTMP-SMZ(15 to 20 mg of trimethoprim per kg, given daily in four divided doses intravenously or by mouth) or pentamidine (3 to 4 mg/kg daily, given intravenously), and therapy is continued for 3 weeks in confirmed cases of PCP. Although some data suggest that TMP-SMZ is more effective than pentamidine, further studies directly comparing the two agents are needed. The frequency and severity of the adverse effects of the two drugs are generally thought to be equivalent. The addition of glucocorticoids (prednisone, 40 mg twice daily, with subsequent tapering of the dose) early in the course of PCP in patients with an arterial P_{O2} of<70 mmHg decreases the need for mechanical ventilation and improves the patient's chances of survival and functional status. Prophylaxis for recurrent PCP must be started at the end of therapy.

Institutionally Acquired Pneumonia Pneumonia acquired in institutions such as nursing homes or hospitals is frequently caused by enteric aerobic gram-negative bacilli, *P. aeruginosa*, or *S. aureus*, with or without oral anaerobes. Again, the selection of empirical antimicrobial therapy should be guided by Gram's staining of sputum (Tables 255-7 and255-8) and knowledge of the prevalent nosocomial pathogens and their current in vitro antimicrobial sensitivity patterns in the institution involved. An aggressive diagnostic approach is needed in some circumstances, especially for the immunocompromised patient (as outlined above).

S. aureus acquired in some institutions is frequently methicillin resistant. Such strains are resistant to all b-lactam antibiotics and may also be resistant to clindamycin, erythromycin, and the fluoroquinolones. Only vancomycin is predictably active against these organisms, and this drug should be added to the empirical regimen when methicillin-resistant organisms may be involved in pneumonia.

When multiantibiotic resistance is a problem, pneumonia due to gram-negative bacilli in the institutionalized patient can be treated initially with a b-lactam active against *P. aeruginosa* (ceftazidime, cefepime, piperacillin/tazobactam, ticarcillin/clavulanate, aztreonam, or imipenem) or with a parenterally administered fluoroquinolone (ciprofloxacin, ofloxacin, gatifloxacin, or levofloxacin). Among the fluoroquinolones, ciprofloxacin remains the most potent antipseudomonal agent. Ticarcillin/clavulanate and piperacillin/tazobactam are preferred over other penicillins with activity against *P. aeruginosa* (e.g., ticarcillin or piperacillin alone), which are not sufficiently active against *Klebsiella pneumoniae*, a relatively common pathogen. However, for infection suspected to be due to *P. aeruginosa*, the higher dose recommended by the package insert is

required; a lower dose contains less piperacillin or ticarcillin than is needed to be effective against this organism. Ampicillin/sulbactam, the other parenterally administeredb-lactam/b-lactamase inhibitor combination, is not active against many nosocomial pathogens, such as *P. aeruginosa*, *Enterobacter* spp., and *Serratia* spp., and therefore is inappropriate as empirical therapy for nosocomial pneumonia.

In seriously ill patients, especially those infected with organisms in which resistance frequently emerges during therapy (e.g., *P. aeruginosa*), use of ab-lactam/aminoglycoside orb-lactam/fluoroquinolone combination is prudent. Combinations of a b-lactam plus an aminoglycoside are used for bactericidal synergy. Combinations of ab-lactam or an aminoglycoside with a fluoroquinolone are not expected to enhance the already-rapid bactericidal activity of the fluoroquinolone alone. However, such combinations are also used to broaden the spectrum of antibacterial activity, to cover the possibility of infection with resistant pathogens, to treat polymicrobial infection, and to prevent the emergence of antimicrobial resistance.

Pneumonia due to possible coinfection with aerobic gram-negative bacilli and anaerobes, as reflected by a polymicrobial flora on Gram's staining of sputum, can usually be treated with any of the following regimens: (1) cefepime or ceftazidime plus metronidazole or clindamycin, (2) aztreonam or a fluoroquinolone plus clindamycin, or (3) imipenem, piperacillin/tazobactam, or ticarcillin/clavulanate. The regimens should include double coverage for *P. aeruginosa* when this organism is suspected (Table 255-8).

The production of chromosomally encoded, inducibleb-lactamases by some aerobic gram-negative bacilli, including *Serratia marcescens*, *Enterobacter cloacae*, *Citrobacter freundii*, *Morganella morganii*, *P. aeruginosa*, and *Acinetobacter calcoaceticus*, has important implications for the treatment of nosocomial pneumonia in institutions where these organisms are common nosocomial pathogens. Antibiotic resistance in these pathogens has been attributed to two related mechanisms: inducible production of chromosomally encoded b-lactamases and selection of mutants that have lost the genes that control expression ofb-lactamase production. The control genes repressb-lactamase production in the absence of ab-lactam agent and allowb-lactamase production in the presence of ab-lactam agent. This group of organisms has a relatively high mutation rate for loss of these control genes, and their loss results in continuous production of large amounts ofb-lactamase (*stable derepression*). The derepressed mutants are resistant to third-generation cephalosporins, aztreonam, and broad-spectrum penicillins. These chromosomally encoded, inducibleb-lactamases are not inhibited by clavulanic acid, tazobactam, or sulbactam.

Selection by theb-lactam antibiotic of the derepressed mutants present in the dense bacterial populations of infected pulmonary tissue at the initiation of antibiotic therapy apparently accounts for the emergence of resistance during therapy, which is especially problematic in severely compromised patients whose defective host defenses are unable to control the growth of a few resistant mutants. The only b-lactam agents that maintain activity against the derepressed mutants are the fourth-generation cephalosporin cefepime and the carbapenem imipenem. The fluoroquinolones and aminoglycosides may also retain activity against these mutants.TMP-SMZmay remain

active against all of these gram-negative bacilli except *P. aeruginosa*, which is inherently resistant to this agent. Some clinicians have questioned the efficacy of aminoglycosides alone for the treatment of gram-negative bacillary pneumonia. The poor clinical efficacy of aminoglycosides has been attributed to the low drug levels attained in bronchial secretions and to a loss of antimicrobial activity due to the relative acidity of purulent secretions, the anaerobic conditions in infected lung, and (in the case of *P. aeruginosa*) the divalent cations calcium and magnesium. The nephrotoxicity and ototoxicity of aminoglycosides frequently lead to underdosing with these agents. These problems are compounded by unpredictable pharmacokinetics that necessitate measurement of serum levels of aminoglycosides. If multiantibiotic-resistant nosocomial organisms are likely to be the pathogens infecting severely compromised patients, reliable empirical agents may be fluoroquinolones, cefepime, and imipenem -- unless resistance to these drugs is also endemic in the institution. Some strains of *K. pneumoniae* and *Escherichia coli* have acquired a plasmid encoding the production of an extended-spectrumb-lactamase that can be detected as in vitro resistance to ceftazidime or aztreonam. The presence of an extended-spectrumb-lactamase confers resistance to all third-generation cephalosporins and aztreonam. Some of these strains may also be resistant to piperacillin/tazobactam and cefepime, and many are also resistant to the fluoroquinolones. The only reliable agents are the carbapenems, such as imipenem. Up-to-date knowledge of the antimicrobial sensitivities of an institution's nosocomial pathogens and use of various preventive practices are mandatory.

Amantadine (200 mg/d for most adults and 100 mg/d for persons >65 years of age) is effective for the prevention of influenza A virus infection in the unimmunized patient during an influenza A outbreak and for the treatment (for 5 to 7 days) of early influenza A virus infection. Ribavirin is effective for respiratory syncytial virus infection. Intravenous acyclovir (5 to 10 mg/kg every 8 h for 7 to 14 days) is appropriate for varicella pneumonia. Treatment of cytomegalovirus pneumonia has yielded unsatisfactory results, but intravenous immunoglobulin combined with ganciclovir may be effective in some instances. Therapy for hantavirus pulmonary syndrome is supportive, and overall mortality has been 55%.

PREVENTION

The prevention of pneumonia involves either (1) decreasing the likelihood of encountering the pathogen or (2) strengthening the host's response once the pathogen is encountered. The first approach can include measures such as hand washing and glove use by persons who care for patients infected with contact-transmitted pathogens (e.g., aerobic gram-negative bacilli); use of face masks or negative-pressure isolation rooms for patients with pneumonia due to pathogens spread by the aerosol route (e.g., *M. tuberculosis*); prompt institution of effective chemotherapy for patients with contagious illnesses; and correction of conditions that facilitate aspiration. The second approach includes the use of chemoprophylaxis or immunization for patients at risk. Chemoprophylaxis may be administered to patients who have encountered or are likely to encounter the pathogen before they become symptomatic (e.g., amantadine during a community outbreak of influenza A, as mentioned above; isoniazid for tuberculosis; orTMP-SMZfor pneumocystosis) or to patients who are likely to have a recurrence following recovery from a symptomatic episode (e.g., TMP-SMZ for pneumocystosis in patients with HIV infection). The prevention of nosocomial pneumonia requires good

infection control practices, judicious use of broad-spectrum antimicrobial agents, and maintenance of patients' gastric acidity -- a major factor that prevents colonization of the gastrointestinal tract by nosocomial gram-negative bacillary pathogens. To prevent stress ulceration, it is preferable to use sucralfate, which maintains gastric acidity, rather than H_2-blocking agents. To prevent ventilator-associated nosocomial pneumonia, the following strategies have been proposed: use of the semirecumbent position, of endotracheal tubes that allow continuous aspiration of secretions accumulating above the cuff, and of heat and moisture exchangers that reduce the formation of condensate within the tubing circuitry. Vaccines (Chaps. 122,138,149,190, and194) are available for immunization against *S. pneumoniae*, *H. influenzae* type b, influenza viruses A and B, and measles virus. Influenza vaccine is strongly recommended for individuals > 55 years old and pneumococcal vaccine for those > 65 years old; these vaccines should be administered to persons of any age who are at risk of adverse consequences of influenza or pneumonia because of underlying conditions. Pneumococcal, *Haemophilus*, and influenza vaccines are recommended for HIV-infected patients who are still capable of responding to a vaccine challenge. The currently available 23-valent pneumococcal vaccine covers 88% of the serotypes causing systemic disease as well as 8% of related serotypes. The increasing prevalence of multiantibiotic resistance among pneumococci makes pneumococcal immunization of high-risk individuals of utmost importance. Immune serum globulin is available for intravenous replacement therapy in those patients with congenital or acquired hypogammaglobulinemia. Some patients who have selective IgG2 subtype deficiency and recurrent sinopulmonary infections and who are immunologically unresponsive to capsular polysaccharide vaccines may nevertheless have an antibody response to the capsular polysaccharide that is covalently linked to a protein, as it is in the conjugate *H. influenzae* type b vaccine and a similar experimental conjugate pneumococcal vaccine.

(Bibliography omitted in Palm version)

Back to Table of Contents

256. BRONCHIECTASIS - *Steven E. Weinberger*

DEFINITION

Bronchiectasis is an abnormal and permanent dilatation of bronchi. It may be either focal, involving airways supplying a limited region of pulmonary parenchyma, or diffuse, involving airways in a more widespread distribution. Although this definition is based on pathologic changes in the bronchi, diagnosis is often suggested by the clinical consequences of chronic or recurrent infection in the dilated airways and the associated secretions that pool within these airways.

PATHOLOGY

The bronchial dilatation of bronchiectasis is associated with destructive and inflammatory changes in the walls of medium-sized airways, often at the level of segmental or subsegmental bronchi. The normal structural components of the wall, including cartilage, muscle, and elastic tissue, are destroyed and may be replaced by fibrous tissue. The dilated airways frequently contain pools of thick, purulent material, while more peripheral airways are often occluded by secretions or obliterated and replaced by fibrous tissue. Additional microscopic features include bronchial and peribronchial inflammation and fibrosis, ulceration of the bronchial wall, squamous metaplasia, and mucous gland hyperplasia. The parenchyma normally supplied by the affected airways is abnormal, containing varying combinations of fibrosis, emphysema, bronchopneumonia, and atelectasis. As a result of the inflammation, vascularity of the bronchial wall increases, with associated enlargement of the bronchial arteries and anastomoses between the bronchial and pulmonary arterial circulations.

Three different patterns of bronchiectasis were described by Reid in 1950. In *cylindrical bronchiectasis* the bronchi appear as uniformly dilated tubes that end abruptly at the point that smaller airways are obstructed by secretions. In *varicose bronchiectasis* the affected bronchi have an irregular or beaded pattern of dilatation resembling varicose veins. In *saccular (cystic) bronchiectasis* the bronchi have a ballooned appearance at the periphery, ending in blind sacs without recognizable bronchial structures distal to the sacs.

ETIOLOGY AND PATHOGENESIS

Bronchiectasis is a consequence of inflammation and destruction of the structural components of the bronchial wall. Infection is the usual cause of the inflammation; microorganisms such as *Pseudomonas aeruginosa* and *Haemophilus influenzae* produce pigments, proteases, and other toxins that injure the respiratory epithelium and impair mucociliary clearance. The host inflammatory response induces epithelial injury, largely as a result of mediators released from neutrophils. As protection against infection is compromised, the dilated airways become more susceptible to colonization and growth of bacteria. Thus, a reinforcing cycle can result, with inflammation producing airway damage, impaired clearance of microorganisms, and further infection, which then completes the cycle by inciting more inflammation.

Infectious Causes Adenovirus and influenza virus are the main viruses that cause

bronchiectasis in association with lower respiratory tract involvement. Virulent bacterial infections, especially with potentially necrotizing organisms such as *Staphylococcus aureus*, *Klebsiella*, and anaerobes, remain important causes of bronchiectasis when antibiotic treatment of a pneumonia is not given or is significantly delayed. Bronchiectasis has been reported in patients with HIV infection, perhaps at least partly due to recurrent bacterial infection. Tuberculosis can produce bronchiectasis by a necrotizing effect on pulmonary parenchyma and airways and indirectly as a consequence of airway obstruction from bronchostenosis or extrinsic compression by lymph nodes. Nontuberculous mycobacteria are frequently cultured from patients with bronchiectasis, often as secondary infections or colonizing organisms. However, it has now also been recognized that these organisms, especially those of the *Mycobacterium avium* complex, can serve as primary pathogens associated with the development and/or progression of bronchiectasis. Mycoplasmal and necrotizing fungal infections are rare causes of bronchiectasis.

Impaired host defense mechanisms are often involved in the predisposition to recurrent infections. The major cause of localized impairment of host defenses is endobronchial obstruction. Bacteria and secretions cannot be cleared adequately from the obstructed airway, which develops recurrent or chronic infection. Slowly growing endobronchial neoplasms such as carcinoid tumors may be associated with bronchiectasis. Foreign-body aspiration is another important cause of endobronchial obstruction, particularly in children. Airway obstruction can also result from bronchostenosis, from impacted secretions, or from extrinsic compression by enlarged lymph nodes.

Generalized impairment of pulmonary defense mechanisms occurs with immunoglobulin deficiency, primary ciliary disorders, or cystic fibrosis. Infections and bronchiectasis are therefore often more diffuse. With panhypogammaglobulinemia, the best described of the immunoglobulin disorders associated with recurrent infection and bronchiectasis, patients often also have a history of sinus or skin infections. Selective deficiency of an IgG subclass, especially IgG2, has also been described in a small number of patients with bronchiectasis.

The primary disorders associated with ciliary dysfunction, termed *primary ciliary dyskinesia*, are responsible for 5 to 10% of cases of bronchiectasis. Numerous defects are encompassed under this category, including structural abnormalities of the dynein arms, radial spokes, and microtubules. The cilia become dyskinetic; their coordinated, propulsive action is diminished, and bacterial clearance is impaired. The clinical effects include recurrent upper and lower respiratory tract infections, such as sinusitis, otitis media, and bronchiectasis. Because normal sperm motility also depends on proper ciliary function, males are generally infertile (Chap. 335). Approximately half of patients with primary ciliary dyskinesia fall into the subgroup of *Kartagener's syndrome*, in which situs inversus accompanies bronchiectasis and sinusitis.

In cystic fibrosis (Chap. 257), the tenacious secretions in the bronchi are associated with impaired bacterial clearance, resulting in colonization and recurrent infection with a variety of organisms, particularly mucoid strains of *P. aeruginosa* but also *S. aureus*, *H. influenzae*, *Escherichia coli*, and *Burkholderia cepacia*.

Noninfectious Causes Some cases of bronchiectasis are associated with exposure to

a toxic substance that incites a severe inflammatory response. Examples include inhalation of a toxic gas such as ammonia or aspiration of acidic gastric contents, though the latter problem is often also complicated by aspiration of bacteria. An immune response in the airway may also trigger inflammation, destructive changes, and bronchial dilatation. This mechanism is presumably responsible at least in part for bronchiectasis with allergic bronchopulmonary aspergillosis (ABPA), which is due to an immune response to *Aspergillus* organisms that have colonized the airway (Chap. 253). Bronchiectasis accompanying ABPA often involves proximal airways and is associated with mucoid impaction. Bronchiectasis also occurs rarely in ulcerative colitis, rheumatoid arthritis, and Sjogren's syndrome, but it is not known whether an immune response triggers airway inflammation in these patients.

In α_1-antitrypsin deficiency, the usual respiratory complication is the early development of panacinar emphysema, but affected individuals may occasionally have bronchiectasis. In the *yellow nail syndrome*, which is due to hypoplastic lymphatics, the triad of lymphedema, pleural effusion, and yellow discoloration of the nails is accompanied by bronchiectasis in approximately 40% of patients.

CLINICAL MANIFESTATIONS

Patients typically present with persistent or recurrent cough and purulent sputum production. Hemoptysis occurs in 50 to 70% of cases and can be due to bleeding from friable, inflamed airway mucosa. More significant, even massive bleeding is often a consequence of bleeding from hypertrophied bronchial arteries.

When a specific infectious episode initiates bronchiectasis, patients may describe a severe pneumonia followed by chronic cough and sputum production. Alternatively, patients without a dramatic initiating event often describe the insidious onset of symptoms. In some cases, patients are either asymptomatic or have a nonproductive cough, often associated with "dry" bronchiectasis in an upper lobe. Dyspnea or wheezing generally reflects either widespread bronchiectasis or underlying chronic obstructive pulmonary disease. With exacerbations of infection, the amount of sputum increases, it becomes more purulent and often more bloody, and patients may become febrile. Such episodes may be due solely to exacerbations of the airway infection, but associated parenchymal infiltrates sometimes reflect an adjacent pneumonia.

Physical examination of the chest overlying an area of bronchiectasis is quite variable. Any combination of crackles, rhonchi, and wheezes may be heard, all of which reflect the damaged airways containing significant secretions. As with other types of chronic intrathoracic infection, clubbing may be present. Patients with severe, diffuse disease, particularly those with chronic hypoxemia, may have associated cor pulmonale and right ventricular failure. Amyloidosis can result from chronic infection and inflammation but is now seldom seen.

RADIOGRAPHIC AND LABORATORY FINDINGS

Though the chest radiograph is important in the evaluation of suspected bronchiectasis, the findings are often nonspecific. At one extreme, the radiograph may be normal with mild disease. Alternatively, patients with saccular bronchiectasis may have prominent

cystic spaces, either with or without air-liquid levels, corresponding to the dilated airways. These may be difficult to distinguish from enlarged airspaces due to bullous emphysema or from regions of honeycombing in patients with severe interstitial lung disease. Other findings are due to dilated airways with thickened walls, which result from peribronchial inflammation. Because of decreased aeration and atelectasis of the associated pulmonary parenchyma, these dilated airways are often crowded together in parallel. When seen longitudinally, the airways appear as "tram tracks"; when seen in cross-section, they produce "ring shadows." Because the dilated airways may be filled with secretions, the lumen may appear dense rather than radiolucent, producing an opaque tubular or branched tubular structure.

Bronchography, which involves coating the airways with a radiopaque, iodinated lipid dye instilled through a catheter or bronchoscope, can provide excellent visualization of bronchiectatic airways. However, this technique has now been replaced by computed tomography (CT), which also provides an excellent view of dilated airways as seen in cross-sectional images (Fig. 256-1). With the advent of high-resolution CT scanning, in which the images are 1.0 to 1.5 mm thick, the sensitivity for detecting bronchiectasis has improved even further. Other features on high-resolution CT scanning can suggest a specific etiology of the bronchiectasis. For example, bronchiectasis of relatively proximal airways suggestsABPA, whereas the presence of multiple small pulmonary nodules (nodular bronchiectasis) suggests infection with *M. avium* complex.

Examination of sputum often reveals an abundance of neutrophils and colonization or infection with a variety of possible organisms. Appropriate staining and culturing of sputum often provide a guide to antibiotic therapy.

Additional evaluation is aimed at diagnosing the cause for the bronchiectasis. When bronchiectasis is focal, fiberoptic bronchoscopy may reveal an underlying endobronchial obstruction. In other cases, upper lobe involvement may be suggestive of either tuberculosis orABPA. With more widespread disease, measurement of sweat chloride levels for cystic fibrosis, structural or functional assessment of nasal or bronchial cilia or sperm for primary ciliary dyskinesia, and quantitative assessment of immunoglobulins may explain recurrent airway infection. In an asthmatic person with proximal bronchiectasis or other historical features to suggest ABPA, skin testing, serology, and sputum culture for *Aspergillus* are helpful in confirming the diagnosis.

Pulmonary function tests may demonstrate airflow obstruction as a consequence of diffuse bronchiectasis or associated chronic obstructive lung disease. Bronchial hyperreactivity, e.g., to methacholine challenge, and some reversibility of the airflow obstruction with inhaled bronchodilators are relatively common.

TREATMENT

Therapy has four major goals: (1) elimination of an identifiable underlying problem; (2) improved clearance of tracheobronchial secretions; (3) control of infection, particularly during acute exacerbations; and (4) reversal of airflow obstruction. Appropriate treatment should be instituted when a treatable cause is found, for example, treatment of hypogammaglobulinemia with immunoglobulin replacement, tuberculosis with antituberculous agents, andABPAwith glucocorticoids.

Secretions are typically copious and thick and contribute to the symptoms. Chest physical therapy with vibration, percussion, and postural drainage frequently helps patients with copious secretions. Mucolytic agents to thin secretions and allow better clearance are controversial. Aerosolized recombinant DNase, which decreases viscosity of sputum by breaking down DNA released from neutrophils, has been shown to improve pulmonary function in cystic fibrosis, but similar benefits have not been found with bronchiectasis due to other etiologies.

Antibiotics have an important role in management. For patients with infrequent exacerbations characterized by an increase in quantity and purulence of the sputum, antibiotics are commonly used only during acute episodes. Although choice of an antibiotic may be guided by Gram's stain and culture of sputum, empiric coverage (e.g., with ampicillin, amoxicillin, trimethoprim-sulfamethoxazole, or cefaclor) is often given initially. When *P. aeruginosa* is present, oral therapy with a quinolone or parenteral therapy with an aminoglycoside or third-generation cephalosporin may be appropriate. In patients with chronic purulent sputum despite short courses of antibiotics, more prolonged courses, e.g., with oral amoxicillin or inhaled aminoglycosides, or intermittent but regular courses of single or rotating antibiotics have been used.

Bronchodilators to improve obstruction and aid clearance of secretions are particularly useful in patients with airway hyperreactivity and reversible airflow obstruction. Although surgical therapy was common in the past, more effective antibiotic and supportive therapy has largely replaced surgery. However, when bronchiectasis is localized and the morbidity is substantial despite adequate medical therapy, surgical resection of the involved region of lung should be considered.

When massive hemoptysis, often originating from the hypertrophied bronchial circulation, does not resolve with conservative therapy, including rest and antibiotics, therapeutic options are either surgical resection or bronchial arterial embolization (Chap. 33). Although resection may be successful if disease is localized, embolization is preferable with widespread disease. In patients with extensive disease, chronic hypoxemia and cor pulmonale may indicate the need for long-term supplemental oxygen. For selected patients who are disabled despite maximal therapy, lung transplantation is a therapeutic option.

(Bibliography omitted in Palm version)

Back to Table of Contents

257. CYSTIC FIBROSIS - *Richard C. Boucher*

Cystic fibrosis (CF) is a monogenetic disorder that presents as a multisystem disease. The first signs and symptoms typically occur in childhood, but about 7% of patients in the United States are diagnosed as adults. Due to improvements in therapy, more than 36% of patients are now adults³18 years of age and 12% are past the age of 30. The median survival is over 32 years for males and 29 years for females with CF. Thus, CF is no longer only a pediatric disease, and internists must be prepared to recognize and treat its many complications. This disease is characterized by chronic airways infection that ultimately leads to bronchiectasis and bronchiolectasis, exocrine pancreatic insufficiency and intestinal dysfunction, abnormal sweat gland function, and urogenital dysfunction.

PATHOGENESIS

GENETIC CONSIDERATIONS

CFis an autosomal recessive disease resulting from mutations in a gene located on chromosome 7. The prevalence of CF varies with the ethnic origin of a population. CF is detected in approximately 1 in 3000 live births in the Caucasian population of North America and northern Europe, 1 in 17,000 live births of African-Americans, and 1 in 90,000 live births of the Asian population of Hawaii. The most common mutation in the CF gene (~70% of CF chromosomes) is a 3-bp deletion that results in an absence of phenylalanine at amino acid position 508 (DF_{508}) of the CF gene protein product, known as the CF transmembrane regulator (CFTR). The large number (>800) of relatively uncommon (<2%) mutations identified in the CF gene makes it difficult to use DNA diagnostic technologies for identifying heterozygotes in populations at large, and no simple physiologic measurements allow heterozygote detection.

CFTRPROTEIN

TheCFTRprotein is a single polypeptide chain containing 1480 amino acids that appears to function both as a cyclic AMP-regulated Cl^- channel and, as its name implies, a regulator of other ion channels. The fully processed form of CFTR is found in the plasma membrane in normal epithelia (Fig. 257-1). Biochemical studies indicate that theDF_{508}mutation leads to improper processing and intracellular degradation of the CFTR protein. Thus, absence of CFTR at appropriate cellular sites is often part of the pathophysiology ofCF. However, other mutations in the CF gene produce CFTR proteins that are fully processed but are nonfunctional or only partially functional at the appropriate cellular sites.

EPITHELIAL DYSFUNCTION

The epithelia affected byCFexhibit different functions in their native state; i.e., some are volume-absorbing (airways and distal intestinal epithelia), some are salt-absorbing but not volume absorbing (sweat duct), and others are volume-secretory (proximal intestine and pancreas). Given this diverse array of native activities, it should not be surprising that CF produces very different effects on patterns of electrolyte and water transport. However, the unifying concept is that all affected tissues express abnormal ion transport

function.

ORGAN-SPECIFIC PATHOPHYSIOLOGY

Lung The diagnostic biophysical hallmark ofCF is the raised transepithelial electric potential difference (PD) detected in airway epithelia. The transepithelial PD reflects components of both the rate of active ion transport and the resistance to ion flow of the superficial epithelium. CF airway epithelia exhibit both raised transport rates (Na^+) and decreased Cl^-permeability (Fig. 257-2). The Cl^- permeability defect reflects at least in part the absence of cyclic AMP-dependent kinase and protein kinase C-regulated Cl^- transport that is mediated by the Cl^-channel functions ofCFTR. An important observation is that there is an alternative Cl^- channel expressed in airway epithelia. This "alternative" Cl^- channel (Cl_{a^-}) is different from CFTR and is regulated by intracellular Ca^{2+}levels. This channel can substitute for CFTR with regard to net Cl^- transport and may be a potential therapeutic target.

Raised Na^+absorption is a routine feature ofCFairway epithelia. Na^+transport abnormalities in CF are not a widespread feature of the CF epithelial phenotype and appear confined to volume-absorbing epithelia. Recent studies demonstrate that the increased Na^+transport reflects the absence ofCFTR's tonic inhibitory regulatory function on Na^+channel activity. It appears that CFTR inhibits Na^+channel activity as a part of its general function to act as a "switch" that coordinates the balance between Na^+absorption and Cl^-secretion.

The central hypothesis ofCFairways pathophysiology has been that an abnormally high rate of Na^+absorption and low rate of Cl^- secretion reduce the salt and water content of mucus and deplete the volume of the perciliary liquid (PCL). Both the thickening of mucins and the depletion of the PCL lead to a failure to clear mucus normally from the airways by either ciliary or airflow-dependent (cough) mechanisms. An alternative hypothesis suggests that the central defect in CF airways is raised salt concentration in secretions that inhibits the function of antimicrobial substances. Direct measurements of salt concentration *in vivo* have, however, provided no evidence that there are differences in salt concentration in CF versus normal airway secretions.

The unique predisposition ofCFairways to chronic infection by *Staphylococcus aureus* and *Pseudomonas aeruginosa* raises the issue that other as yet undefined abnormalities in airway surface liquids also may contribute to the failure of lung defense. However, it may be that *Pseudomonas* is selected by its propensity to grow in biofilm colonies on the surfaces of thickened, retained mucus plaques in CF airways.

Gastrointestinal Tract The gastrointestinal effects ofCF are diverse. In the exocrine pancreas, the absence of theCFTRCl^- channel in the apical membrane of pancreatic ductal epithelia limits the function of an apical membrane Cl^--HCO_3^-exchanger to secrete $HCO_3^-$and Na^+(by a passive process) into the duct. The failure to secrete Na^+-$HCO_3^-$and water leads to retention of enzymes in the pancreas and ultimately destruction of virtually all pancreatic tissue. The CF intestinal epithelium, because of the lack of Cl^- and water secretion, fails to flush the secreted mucins and other macromolecules from intestinal crypts. The diminished CFTR-mediated secretion of liquid may be exacerbated by excessive absorption of liquid in the distal intestine,

reflecting abnormalities of CFTR-mediated regulation of Na^+absorption (both mediated by Na^+channels and possibly other Na^+transporters, e.g., Na^+-H^+exchangers). Both dysfunctions lead to dessicated intraluminal contents and obstruction of both the small and large intestines. In the hepatobiliary system, defective hepatic ductal Cl^- and water secretion causes retention of biliary secretions and focal biliary cirrhosis and bile duct proliferation in approximately 25 to 30% of patients with CF. The inability of the CF gallbladder epithelium to secrete salt and water can lead to both chronic cholecystitis and cholelithiasis.

Sweat Gland Patients with CFsecrete nearly normal volumes of sweat in the sweat acinus. However, they are not able to absorb NaCl from sweat as it moves through the sweat duct due to the inability to absorb Cl^- across the ductal epithelial cells.

CLINICAL FEATURES

Most patients withCFpresent with signs and symptoms of the disease in childhood. Approximately 15% of patients present within the first 24 h of life with gastrointestinal obstruction, termed *meconium ileus*. Other common presentations within the first year or two of life include respiratory tract symptoms, most prominently cough and/or recurrent pulmonary infiltrates, and failure to thrive. A significant proportion of patients (~7%), however, are diagnosed after age 18.

RESPIRATORY TRACT

Upper respiratory tract disease is almost universal in patients withCF. Chronic sinusitis is common in childhood and leads to nasal obstruction and rhinorrhea. The occurrence of nasal polyps approaches 25% and often requires surgery.

In the lower respiratory tract, the first symptom ofCF is cough. With time, the cough becomes persistent and produces viscous, purulent, often greenish colored sputum. Inevitably, periods of clinical stability are interrupted by "exacerbations," defined by increased cough, weight loss, increased sputum volume, and decrements in pulmonary function. These exacerbations require aggressive therapy, including frequent postural drainage and oral antibiotics, and often intravenous antibiotics (see below), with the goal being recovery of lung function. Over the course of years, the exacerbations become more frequent and the recovery of lost lung function incomplete, leading to respiratory failure.

Patients withCFexhibit a characteristic sputum microbiology. *Haemophilus influenzae* and *S. aureus* are often the first organisms recovered from samples of lung secretions in newly diagnosed patients with CF. *P. aeruginosa* is typically cultured from lower respiratory tract secretions thereafter. After repetitive antibiotic exposure, *P. aeruginosa*, often in a mucoid form, is usually the predominant organism recovered from sputum and may be present as several strains with different antibiotic sensitivities. *Burkholderia* (formerly *Pseudomonas*) *cepacia* has been recovered from CF sputum and is pathogenic. Patient-to-patient spread of certain strains of this organism indicates that infection control in the hospital should be practiced. Other gram-negative rods recovered from CF sputum include *Xanthomonas zylosoxida* and *P. gladioli*, and occasionally, mucoid forms of *Proteus*, *Escherichia coli*, and *Klebsiella*. Up to 50% of

patients with CF have *Aspergillus fumigatus* in their sputum, and up to 10% of these patients exhibit the syndrome of allergic bronchopulmonary aspergillosis. *Mycobacterium tuberculosis* is rare in patients with CF. However, 10 to 20% of adult patients with CF have sputum cultures positive for nontuberculous mycobacteria, and in some patients these microorganisms are associated with disease.

The first lung function abnormalities observed in children withCF, increased ratios of residual volume to total lung capacity, suggest that small airways disease is the first functional lung abnormality in CF. As the disease progresses, both reversible and irreversible changes in forced vital capacity and forced expiratory volume in 1 s are noted. The reversible component reflects the accumulation of intraluminal secretions and/or airway reactivity, which occurs in 40 to 60% of patients with CF. The irreversible component reflects chronic destruction of the airway wall and bronchiolitis.

The earliest chest x-ray change in CFlungs is hyperinflation, reflecting small airways obstruction. Later, signs of luminal mucus impaction, bronchial cuffing, and finally, bronchiectasis, e.g., ring shadows, are noted. For reasons that are still unknown, the right upper lobe displays the earliest and most severe changes. Neither CT nor MRI scanning is routinely performed on patients with CF.

CFpulmonary disease is associated with many intermittent complications. Pneumothorax is common (>10% of patients). The production of small amounts of blood in sputum is common in CF patients with advanced pulmonary disease and appears to be associated with lung infection. Massive hemoptysis is life-threatening and difficult to localize bronchoscopically. With advanced lung disease, digital clubbing becomes evident in virtually all patients with CF. As late events, respiratory failure and cor pulmonale are prominent features of CF.

GASTROINTESTINAL TRACT

The syndrome of meconium ileus in infants presents with abdominal distention, failure to pass stool, and emesis. The abdominal flat plate can be diagnostic with small intestinal air fluid levels, a granular appearance representing meconium, and a small colon. In children and young adults, a syndrome termed *meconium ileus equivalent* or distal intestinal obstruction occurs. The syndrome presents with right lower quadrant pain, loss of appetite, occasional emesis, and often a palpable mass. The syndrome can be confused with appendicitis, which occurs frequently in patients withCF. The characteristic intestinal abnormalities are complicated by exocrine pancreatic insufficiency in more than 90% of patients with CF. Insufficient pancreatic enzyme release yields the typical pattern of protein and fat malabsorption, with frequent, bulky, foul-smelling stools. Signs and symptoms of malabsorption of fat-soluble vitamins, including vitamins E and K, are also noted. Pancreatic beta cells are typically spared, but function decreases with age, causing hyperglycemia and increasing requirements for insulin in older patients with CF.

GENITOURINARY SYSTEM

Late onset of puberty is common in both males and females withCF. The delayed maturational pattern is likely secondary to the effects of chronic lung disease and

inadequate nutrition on reproductive endocrine function. More than 95% of male patients with CF are azoospermic, reflecting obliteration of the vas deferens that probably reflects defective liquid secretion. Twenty percent of women with CF are infertile due to effects of chronic lung disease on the menstrual cycle; thick, tenacious cervical mucus that blocks sperm migration; and possibly fallopian tube/uterine wall abnormalities in liquid transport. More than 90% of completed pregnancies produce viable infants, and women with CF are generally able to breast-feed infants normally.

DIAGNOSIS

Because of the large number ofCFmutations, DNA analysis is not used for primary diagnosis. The primary diagnosis of CF rests on a combination of clinical criteria and analyses of sweat Cl^- values. The values for the Na^+ and Cl^- concentration in sweat vary with age, but typically in adults a Cl^- concentration of>70 mEq/L discriminates between patients with CF and patients with other lung diseases.

DNA analyses are being performed increasingly in patients withCF. Comprehensive genotype-phenotype relationships have not yet been established sufficiently for prognosis. A relationship betweenDF_{508}homozygosity and pancreatic insufficiency has been established, but no predictive relationship holds forDF_{508}homozygosity and lung disease.

Between 1 and 2% of patients with the clinical syndrome ofCF have normal sweat Cl^-values. In most of these patients, the nasal transepithelialPD is raised into the diagnostic range for CF, and sweat acini do not secrete in response to injected beta-adrenergic agonists. A single mutation of theCFTRgene, 3849 + 10 kb C ® T, is associated with approximately 50% of CF patients with normal sweat Cl^-values.

TREATMENT

The major objectives of therapy forCF are to promote clearance of secretions and control infection in the lung, provide adequate nutrition, and prevent intestinal obstruction. Ultimately, gene therapy may become the treatment of choice.

Lung Disease The principal techniques for clearing pulmonary secretions are breathing exercises, flutter valves, and chest percussion. Regular use of these maneuvers is effective in preserving lung function. There is increasing interest in the use of hypertonic saline (3 to 7%) aerosols to augment the clearance of secretions.

More than 95% of patients withCF die of complications resulting from lung infection. Antibiotics are the principal agents available for treating lung infection, and their use should be guided by sputum culture results. Early intervention with antibiotics is useful, and long courses of treatment are the rule. Because of increased total-body clearance and volume of distribution of antibiotics in patients with CF, the required doses are higher for patients with CF than for patients with similar chest infections who do not have CF.

Increased cough and mucus production are treated with antibiotics given orally. Typical oral agents used to treat *Staphylococcus* include a semisynthetic penicillin or a

cephalosporin. Oral ciprofloxacin may reduce pseudomonal bacterial counts and control symptoms. However, its clinical usefulness may be limited by rapid emergence of resistant organisms, and accordingly, courses should be intermittent (2 to 3 weeks) and not chronic. More severe exacerbations, or exacerbations associated with bacteria resistant to oral antibiotics, require intravenous antibiotics. Traditionally, intravenous therapy has been given in the hospital, but outpatient intravenous antibiotic administration has gained widespread acceptance. Usually, two drugs, often one of them an aminoglycoside, are used to treat *P. aeruginosa* to hinder emergence of resistant organisms. Drug dosage should be monitored so that levels for gentamicin or tobramycin peak at ranges of ~10 ug/mL and exhibit troughs of<2 ug/mL. Usually, a cephalosporin, e.g., ceftazadime, and/or a penicillin derivative is used as the second drug. Antibiotics directed at *Staphylococcus* and/or *H. influenzae* are added depending on the results of the culture. Aerosolization of antibiotics also may have an important role in treatingCF lung infection. Large doses of aminoglycosides, e.g., 600 mg tobramycin twice daily, via aerosol may be effective at delaying exacerbations. Aerosol administration also permits the use of other drugs, e.g., colistin, that are relatively ineffective by the intravenous route.

A number of pharmacologic agents for promoting mucus clearance are in use. *N*-acetyl-cysteine, which solubilizes mucus glycoproteins, has not been shown to have clinically significant effects on mucus clearance and/or lung function. Recombinant human DNAse, however, degrades the concentrated DNA inCFsputum, decreases sputum viscosity, and increases airflow during short-term administration. Long-term (6 months) DNAse treatment increases the time between pulmonary exacerbations. Most patients receive a therapeutic trial of DNAse to test for efficacy, and a sizeable minority appear to demonstrate persistent objective benefits. Clinical trials of experimental drugs aimed at restoring salt and water content of secretions are underway. The most promising may be long-acting nucleotide (UTP)-based compounds that appear active in inducing liquid secretion in CF airways.

Inhaledb-adrenergic agonists can be useful to control airways constriction. They achieve a short-term increase in airflow, but long-term benefit has not been shown. Inhaled anticholinergics provide an alternative. Oral steroids are not first-line agents for controlling airways constriction and are of no use in improving the nonreversible component of lung function. Steroids may be useful for treating allergic bronchopulmonary aspergillosis.

The chronic damage to airway walls reflects to some extent the destructive activities of inflammatory enzymes generated in part by inflammatory cells. To date, specific therapies with antiproteases have not been successfully developed. However, a subset of adolescents withCFappears to benefit from long-term, high-dose non-steroidal (ibuprofen) therapy.

A number of pulmonary complications require acute interventions. Atelectasis is best treated with chest physiotherapy and antibiotic therapy. Pneumothoraces involving 10% or less of the lung can be observed without intervention. The use of chest tubes to expand collapsed, diseased lung often requires long periods of time, and sclerosing agents should be used with caution because of possible limitations for subsequent lung transplantation. Small-volume hemoptysis requires no specific therapy other than

treatment of lung infection and assessment of coagulation and vitamin K status. If massive hemoptysis occurs, bronchial artery embolization can be successful. The most ominous complications ofCF are respiratory failure and cor pulmonale. The most effective conventional therapy for these conditions is vigorous medical management of the lung disease and O_2supplementation. Noninvasive positive pressure ventilation through a face mask may be an effective adjunctive therapy. Ultimately, the only effective treatment for respiratory failure in CF is lung transplantation (Chap. 267). The 2-year survival for lung transplantation exceeds 60%, and deaths in transplant patients result principally from graft rejection, often involving obliterative bronchiolitis. The transplanted lungs do not develop a CF-specific phenotype.

Gastrointestinal Disease Maintenance of adequate nutrition is critical for the health of the patient withCF. Most (>90%) of patients with CF benefit from pancreatic enzyme replacement. Capsules generally contain between 4000 and 29,000 units of lipase. The dose of enzymes (typically no more than 20,000 units/kg per meal) should be adjusted on the basis of weight gain, abdominal symptomatology, and character of stools. Replacement of fat-soluble vitamins, particularly vitamins E and K, is usually required. Hyperglycemia most often becomes manifest in the adult and typically requires insulin treatment.

For treatment of acute obstruction due to meconium ileus equivalent, megalodiatrizoate or other hypertonic radiocontrast materials delivered by enema to the terminal ileum are utilized. For control of symptoms, adjustment of pancreatic enzymes and the supplementation of intake by salt solutions containing osmotically active agents, e.g., propyleneglycol or lactulose, are utilized. Persistent symptoms may indicate a diagnosis of gastrointestinal malignancy, which is increased in incidence in patients withCF. Hepatic and gallbladder complications are treated as for patients without CF. End-stage liver disease can be treated by transplantation, which has a 2-year survival rate exceeding 50%.

Psychosocial Factors CFimposes a tremendous burden on patients. Health insurance, career options, family planning, and life expectancy become major issues. Thus, assisting patients with the psychosocial adjustments required by CF is critical.

(Bibliography omitted in Palm version)

Back to Table of Contents

258. CHRONIC BRONCHITIS, EMPHYSEMA, AND AIRWAYS OBSTRUCTION - *Eric G. Honig*, *Roland H. Ingram, Jr.*

DEFINITION

Chronic obstructive pulmonary disease (COPD) is the name of a group of chronic and slowly progressive respiratory disorders characterized by reduced maximal expiratory flow during forced exhalation. Most of the airflow obstruction is fixed, but a variable degree of reversibility and bronchial hyperreactivity may be seen. COPD may coexist with asthma and, when abnormal airway reactivity is present, differentiation between these disorders can be challenging. COPD comprises emphysema and chronic bronchitis, two distinct processes, although most often present in combination. The definition excludes other causes of chronic airflow obstruction such as cystic fibrosis (Chap. 257), bronchiolitis obliterans (Chap. 259), and bronchiectasis (Chap. 256).

Emphysema is defined anatomically as a permanent and destructive enlargement of airspaces distal to the terminal bronchioles without obvious fibrosis and with loss of normal architecture. *Chronic bronchitis* is defined clinically as the presence of a cough productive of sputum not attributable to other causes on most days for at least 3 months over 2 consecutive years. Chronic bronchitis may be present in the absence of airflow limitation, butCOPDalways involves clinically significant airflow limitation.

EPIDEMIOLOGY

COPDis a common medical problem affecting an estimated 16 million Americans. Males are more frequently affected than females, and Caucasians more frequently than African Americans. There is a higher prevalence of COPD among persons with a lower socioeconomic status and in those with a history of low birth weight. COPD is the fourth leading cause of death in the United States and is the only one of the 10 leading causes of death for which mortality rates are still rising. Prevalence peaks in the seventh and eighth decades, then levels off, largely due to mortality.

DISEASE MECHANISMS

PATHOGENESIS

COPDevolves from an inflammatory process involving the airways and distal airspaces. Increased activity of oxidants combined with decreased activity of antioxidants, termed *oxidative stress*, have been implicated in the development of inflammation and COPD. Cigarette smoke produces high concentrations of oxygen free radicals including superoxide, hydrogen peroxide, and hypochlorous acid. Cigarette smoke is an independent source of Fe^{2+}, releases $Fe^{2+}$from ferritin, and catalyzes the formation of the highly active hydroxyl radical from $O_2^-$and $H_2O_2$by eosinophils, neutrophils, and alveolar macrophages. Cigarette tar contains nitric oxide and induces nitric oxide synthase. In the presence of oxidants, NO is metabolized to cytotoxic peroxynitrates. In order for elastase to degrade elastin,$a_1$antitrypsin (a_1AT) must be inactivated. Cigarette smoke, oxidants, activated neutrophils, and type II alveolar pneumocytes are all capable of inactivating a_1AT as well as matrix metalloproteinase inhibitors. Oxidant stress is also capable of inducing mucus hypersecretion. Cigarette smoke also acts as a

chemoattractant and upregulates adhesion molecules. Smoke increases neutrophil transit time through the pulmonary circulation, increases adhesion, and decreases deformability. Smoke and elastase both increase the expression of the proinflammatory nuclear transcription factor kB (NfkB) as well as interleukin 8, a chemokine found to be elevated in COPD patients, that recruits neutrophils, basophils, eosinophils, and T lymphocytes.

The submucosa of the small airway in patients withCOPD has increased numbers of CD8 lymphocytes and eosinophils, macrophages, and mast cells. Neutrophils are increased in smokers, but their numbers do not correlate with the presence of airflow obstruction. Patients with chronic airflow obstruction show higher levels of myeloperoxidase and eosinophilic cationic protein than do patients with normal airflow. Macrophages and mast cells produce transforming growth factor b(TGF-b), a peptide related to fibrogenesis. Patients with chronic airflow obstruction show a twofold elevation of TGF-b in lavage liquid; the amount of TGF-b shows a significant negative correlation with FEV_1(the forced expiratory volume in 1 s). Smoke also leads to lipid peroxidation and to DNA damage. Widespread point mutations of the p53 gene locus have been identified in patients with lung cancer and precancerous dysplasia. These may predispose to the development of lung cancer.

RISK FACTORS

COPDis characterized by a reducedFEV$_1$and an accelerated rate of decline of FEV_1. The reduction in FEV_1 can occur by any of three pathways: (1) impaired childhood growth and development, with a lower peak in early adulthood and a normal rate of decline with aging (e.g., early childhood infection and passive smoke exposure); (2) normal growth and development with a premature peak but normal subsequent decline (e.g., asthma and passive smoking); and (3) normal growth and development and peak with accelerated decline (e.g., active smoking and, to a lesser degree, environmental exposures).

Smoking Cigarette smoking is the most commonly identified correlate with both chronic bronchitis during life and extent of emphysema at postmortem. The prevalence ofCOPDshows a dose-response relationship with the number of pack-years of tobacco consumed. Some 90% of all COPD patients are current or former tobacco smokers. Experimental studies have shown that prolonged cigarette smoking impairs respiratory epithelial ciliary movement, inhibits function of alveolar macrophages, and leads to hypertrophy and hyperplasia of mucus-secreting glands; massive exposure in dogs can produce emphysematous changes. Cigarette smoke also inhibits antiproteases and causes polymorphonuclear leukocytes to release proteolytic enzymes acutely. Cigarette smoke can produce an acute increase in airways resistance due to vagally mediated smooth-muscle constriction by stimulating submucosal irritant receptors. Increased airways responsiveness is associated with more rapid progression in patients with chronic airways obstruction. Obstruction of small airways is the earliest demonstrable mechanical defect in young cigarette smokers and may disappear completely after cessation of smoking.

Although smoking cessation does not result in complete reversal of more pronounced obstruction, there is a significant slowing of the decline in lung function in all smokers

who give up cigarettes. Passive exposure to tobacco smoke correlates with respiratory symptoms such as cough, wheeze, and sputum production. Not only is cigarette smoking the most common single factor leading to chronic airways obstruction, it also adds to the effects of every other contributory factor to be discussed below.

Air Pollution The incidence and mortality rates of both chronic bronchitis and emphysema may be higher in heavily industrialized urban areas. Exacerbations of bronchitis are clearly related to periods of heavy pollution with sulfur dioxide (SO_2) and particulate matter. While nitrogen dioxide (NO_2) can produce small-airways obstruction (bronchiolitis) in experimental animals exposed to high concentrations, there are no data convincingly implicating NO_2, at even the highest pollutant levels, in the pathogenesis or worsening of airways obstruction in humans (Chap. 254).

Occupation Chronic bronchitis is more prevalent in workers who engage in occupations exposing them to either inorganic or organic dusts or to noxious gases. Epidemiologic surveys have succeeded in demonstrating an accelerated decline in lung function in many such workers -- e.g., workers in plastics plants exposed to toluene diisocyanate, and carding room workers in cotton mills (Chap. 254) -- suggesting that their occupational exposure contributes to their future disability.

Infection Morbidity, mortality, and frequency of acute respiratory illnesses are higher in patients with chronic bronchitis. Many attempts have been made to relate these illnesses to infection with viruses, mycoplasmas, and bacteria. However, only the rhinovirus is found more often during exacerbations; that is to say, pathogenic bacteria, mycoplasmas, and viruses other than rhinovirus are found just as often between as during exacerbations. Epidemiologic studies, however, implicate acute respiratory illness as one of the major factors associated with the etiology as well as the progression of chronic airways obstruction. Cigarette smokers may either transitorily develop or worsen small-airways obstruction in association with even mild viral respiratory infections. There is also some evidence that severe viral pneumonia early in life may lead to chronic obstruction, predominantly in small airways.

GENETIC CONSIDERATIONS

Despite the strong etiologic association between smoking and COPD, only 15 to 20% of smokers lose FEV_1 at a rate fast enough to manifest COPD. Epidemiologic evidence of familial clustering of COPD cases is strong and repeated, suggesting that susceptibility to the effects of tobacco smoke has genetic determinants. Twin studies show that even after controlling for active and passive smoking, FEV_1 correlated more closely in monozygotic than dizygotic twins and more than in other family members with a lesser percentage of shared genotype. In first-degree relatives of a cohort of COPD patients with normal α_1AT levels, FEV_1 was reduced compared to controls but only among current or ex-smokers. Smoking and nonsmoking relatives of control subjects both had normal FEV_1. These data suggest genetic risk factors that are expressed in response to smoking.

α_1Antitrypsin Deficiency Thus far, deficiency of α_1AT is the only genetic abnormality that has been specifically linked to COPD. α_1AT is a 394-amino acid serine proteinase inhibitor whose synthesis is governed by a 12.2-kB 7-exon gene located at

14q32.1.a_1AT synthesis is expressed primarily in the liver and to a lesser degree in neutrophils and monocytes. Hepatica_1AT escapes into the general circulation, where it counteracts neutrophil elastase. Normal levels ofa_1AT are 20 to 48 umol/L; levels above 11 umol/L (35% of normal) are considered protective. There are 75 known alleles ofa_1AT, which are inherited in an autosomal codominant manner and are generally classified as normal (MM), deficient, null, or dysfunctional. The most common deficient allele, termed ZZ (or Pi_{ZZ}phenotype), results from a single amino acid substitution$_{342}$Glu® Lys, which causes spontaneous polymerization of the polypeptide, markedly impeding its release into the circulation from the liver. What does escape is vulnerable to oxidation and spontaneous polymerization, further impeding its function. The retained material is associated with hepatic cirrhosis (Chap. 299), while diminished circulating levels (2.5 to 7 umol/L, averaging 16% of normal) lead to antiprotease deficiency. Pi_{ZZ}, the most common disease-relateda_1AT abnormality, occurs in 1:2000 to 1:7000 persons of European descent and is rare in those of Oriental and African lineage. Pi_{SS}phenotypes are associated witha_1AT levels of 15 to 33 umol (mean 52% of normal). Pi_{null}have no detectable antiprotease levels. Heterozygotes have intermediate levels of antiprotease.

Clinically significant deficiency ofa_1AT, with levels below 11 umol/L, has been associated with homozygous Pi_{ZZ}, $Pi_{nullnull}$, or $Pi_{nullZ}$and the premature development of severe emphysema, chronic bronchitis, or bronchiectasis.a_1AT deficiency accounts for 2% of observed cases of emphysema. Rare below age 25, the disease usually presents as dyspnea and cough in patients in their fourth decade. Although not a true population-based study, a large national registry of 1129 severea_1AT-deficiency cases indicated that the typical patient was in the mid-forties, with an$FEV_1$and a pulmonary diffusing capacity at or below 50% of the predicted levels. Most had exertional dyspnea and wheezing, but fewer than half reported a chronic cough. Nearly 80% had a positive family history of lung disease, and 25% reported a positive family history for liver disease. The average rate of decline of FEV_1 is reported to be 100 to 130 mL per year for smokers and 50 to 80 mL per year for ex-smokers or lifetime nonsmokers witha_1AT deficiency.

Pathologically, panacinar emphysema predominates, and radiographically, changes are more marked in the lower lobes. It is becoming increasingly apparent that tobacco smoking is an extremely important cofactor for the development of disease ina_1AT-deficient individuals. Only a few lifetime nonsmokers with Pi_{ZZ}develop emphysema. Most never have symptoms, have a normal rate of decline ofFEV_1, and live a normal life span. Many cases are discovered only as a consequence of family screening of emphysema patients. Because the total number of Pi_{ZZ}individuals is unknown, the risk of disease for smokers is difficult to ascertain accurately. The risk of disease is lower still for heterozygotes with one M or S allele. Smoking is again an important cofactor.

PATHOLOGY

The pathologic changes ofCOPDinvolve large and small airways and the terminal respiratory unit. Airway narrowing is seen in large and small airways and is caused by changes in their normal constituents in response to persistent inflammation.

The airway epithelium is characterized by squamous metaplasia, atrophy of ciliated cells, and hypertrophy of mucus glands. The remodeled epithelium actively produces cytokines that amplify and sustain the inflammatory process. The small airways are the major site of airflow limitation. Small airways show a variety of lesions narrowing their lumina, including goblet cell hyperplasia, mucosal and submucosal inflammatory cells, edema, peribronchial fibrosis, intraluminal mucus plugs, and increased smooth muscle. CD8+ T lymphocytes and B lymphocytes characterize the inflammatory infiltrate. The marked thickening of the subepithelial lamina reticularis, characteristic of asthma, is absent inCOPD.

In the central airways, subepithelial inflammation is present with increased numbers of eosinophils and CD8+ T lymphocytes. Unlike asthma, the eosinophils are not activated and do not degranulate. Neutrophils are present in the epithelium but not in the subepithelial layers. In larger cartilaginous airways, chronic bronchitis is associated with hypertrophy of submucosal mucus-producing glands. Quantitation of this anatomic change, known as the *Reid index*, is based on the ratio of the thickness of the submucosal glands to that of the bronchial wall. In persons without a history of chronic bronchitis, the mean ratio is 0.44 ± 0.09, whereas in those with such a history, the mean ratio is 0.52 ± 0.08. Although a low index is rarely associated with symptoms and a high index is commonly associated with symptoms during life, there is a great deal of overlap. Therefore, many persons will have morphologic changes in large airways without having had chronic bronchitis.

Emphysema begins as an increase in the number and size of alveolar fenestrae and results in the eventual destruction of alveolar septae and their attachments to terminal and respiratory bronchioles. Emphysema is classified according to the pattern of involvement of the gas-exchanging units (acini) of the lung distal to the terminal bronchiole. With *centriacinar emphysema*, the distention and destruction are mainly limited to the respiratory bronchioles with relatively less change peripherally in the acinus. Because of the large functional reserve in the lung, many units must be involved in order for overall dysfunction to be detectable. The centrally destroyed regions of the acinus have a high ventilation/perfusion ratio because the capillaries are missing, yet ventilation continues. This results in a deficit of perfusion relative to ventilation, while the peripheral portions of the acinus have crowded and small alveoli with intact, perfused capillaries giving a low ventilation/perfusion ratio. This results in a deficit of ventilation relative to blood flow, giving a high alveolar-arterial P_{O_2}difference (PA_{O_2}-Pa_{O_2}) (Chap. 250).

During normal aging, airspaces enlarge and alveolar ducts increase in diameter. These changes are extremely common in lungs from persons over age 50 and may be misidentified as emphysema.

Panacinar emphysema involves both the central and peripheral portions of the acinus, which results, if the process is extensive, in a reduction of the alveolar-capillary gas exchange surface and loss of elastic recoil properties. When emphysema is severe, it may be difficult to distinguish between the two types, which most often coexist in the same lung.

PATHOPHYSIOLOGY

Airflow Limitation Although both chronic bronchitis and emphysema can exist without evidence of obstruction, by the time a patient begins to experience dyspnea as a result of these processes, obstruction is always demonstrable. Airflow limitation and increased airways resistance may be caused by loss of elastic recoil driving passive exhalation due to emphysema, by increased collapsibility of small airways through loss of radial traction on airways, or to increased resistance due to intrinsic narrowing of small airways.

In addition to providing radial support to airways during quiet breathing, the elastic recoil properties of the lung serve as a major determinant of maximal expiratory flow rates. The static recoil pressure of the lung is the difference between alveolar and intrapleural pressure. During forced exhalations, when alveolar and intrapleural pressures are high, there are points in the airway at which bronchial pressure equals pleural pressure. Flow does not increase with higher pleural pressure after these points become fixed, so that the effective driving pressure between alveoli and such points is the elastic recoil pressure of the lung (Fig. 258-1). Hence maximal expiratory flow rates represent a complex and dynamic interplay among airways caliber, elastic recoil pressures, and collapsibility of airways. Correlative studies of structure and function suggest that small-airway narrowing is the most important correlate of airflow obstruction, followed by loss of elastic recoil. Collapsibility is probably a less important factor. As a direct consequence of the altered pressure-airflow relationships, the work of breathing is increased in bronchitis and emphysema. Since flow-resistive work is flow rate-dependent, there is a disproportionate increase in the work of breathing when ventilation must be increased, as in exercise.

Hyperinflation The designated subdivisions of the lung volume outlined inChap. 250 are abnormal to varying degrees in both bronchitis and emphysema. The residual volume and functional residual capacity (FRC) are almost always higher than normal. Since the normal FRC is the volume at which the inward recoil of the lung is balanced by the outward recoil of the chest wall, loss of elastic recoil of the lung results in a higher FRC. In addition, prolongation of expiration in association with obstruction would lead to a dynamic increase in FRC (dynamic hyperinflation) if inspiration is initiated before the respiratory system reaches its static balance point. Dynamic hyperinflation contributes additionally to the discomfort associated with airflow obstruction by flattening the diaphragm and placing it at a mechanical disadvantage due to shortened diaphragmatic fiber length and a perpendicular insertion with the lower ribs. The exertional increase in end-expiratory lung volume and consequent decrease in inspiratory capacity have been strongly associated with the degree of dyspnea. Elevations of total lung capacity (TLC) are frequent. The exact cause is uncertain, but increases in total lung capacity are often found in association with decreases in the elastic recoil of the lung. Although the vital capacity is frequently reduced, significant airways obstruction can be present with a normal to near-normal vital capacity.

Impaired Gas Exchange Maldistribution of inspired gas and blood flow is always present to some extent. When the mismatching is severe, impairment of gas exchange is reflected in abnormalities of arterial blood gases. Small-airway narrowing causes a decrease in ventilation of their distal alveolar acini. When alveolar capillaries remain intact, this results in mismatching of ventilation and blood flow, reduced

ventilation-perfusion ratios, and mild to moderate hypoxemia. With emphysema, destruction of alveolar walls may decrease alveolar capillary perfusion as well, better preserving ventilation-perfusion matching, and Pa_{O_2}. Shunt hypoxemia is unusual. There are regions of the lung with a deficit of perfusion in relation to ventilation that increase the wasted ventilation ratio (i.e., V_d/V_t; Chap. 250). At a normal resting CO_2 production, the net effective alveolar ventilation, as reflected by the arterial P_{CO_2}, may be excessive, normal, or insufficient, depending on the relationship of the overall minute volume to the wasted ventilation ratio.

The severity of gas exchange disturbances and, in large part, the clinical manifestations depend on the ventilatory response to the disordered lung function. Some patients, at the cost of extremely high effort of breathing and chronic dyspnea, maintain a strikingly increased minute volume, which results both in a normal to low arterial P_{CO_2}, despite the high V_d/V_t, and a relatively high arterial P_{O_2}, despite the high difference, PA_{O_2}-Pa_{O_2}. Other patients with only modest increases in effort of breathing and less dyspnea maintain a normal to only moderately elevated minute volume at the cost of accepting a high arterial P_{CO_2} and a severely depressed arterial P_{O_2}.

Factors that account for clear differences in ventilatory responses among patients have been studied and debated for years. The bulk of available evidence suggests that those patients who maintain relatively normal or low arterial P_{CO_2} levels are those with an increased ventilatory drive relative to their blood gas values, and those who chronically maintain high arterial P_{CO_2} and lower P_{O_2} levels have a diminished ventilatory drive in relation to their more severely deranged blood gas values. It is not at all certain whether individual differences are accounted for by variations in peripheral or central chemoreceptor sensitivity or through other afferent pathways.

Pulmonary Circulation The pulmonary circulation malfunctions not only in terms of regional distribution of blood flow but also in terms of abnormal overall pressure-flow relationships. In advanced disease, there is often mild to severe pulmonary hypertension at rest, with further increases disproportionate to cardiac output elevations during exercise. A reduction in the total cross-sectional area of the pulmonary vascular bed can be attributed to thickening of medium and large muscular pulmonary arteries, to enhanced contraction of vascular smooth muscle in pulmonary arteries and arterioles, as well as to destruction of alveolar septa with loss of capillaries. Rarely does loss of capillaries alone lead to severe pulmonary hypertension with cor pulmonale, except as a near-terminal event (Chap. 237). Of more importance is the constriction of pulmonary vessels in response to alveolar hypoxia. The pulmonary arteries of patients with severe hypoxemia COPD have been shown to exhibit increased contractility and impaired relaxation in response to pharmacologic stimuli in vitro. These differences between the pulmonary arteries of COPD patients and normal individuals are abolished by inhibition of NO synthase, suggesting that patients develop an endothelial defect in NO synthesis. The constriction is somewhat reversible by an increase in alveolar P_{O_2} with therapy.

There is a synergism between hypoxia and acidosis that assumes importance during episodes of acute or chronic respiratory insufficiency. Chronic hypoxia, especially in concert with carboxyhemoglobinemia, often seen with heavy cigarette smoking, leads not only to pulmonary vascular constriction but also to secondary erythrocytosis. The latter, although not proved to be a significant contributor to pulmonary hypertension,

could add to pulmonary vascular resistance. As discussed inChap. 237, chronic afterload on the right ventricle leads to hypertrophy and, in association with disordered blood gases, ultimately to failure. Hypoventilation may occur during rapid eye movement sleep and lead to desaturation, which may be severe. Repeated desaturation may cause pulmonary hypertension.

Renal and Hormonal Dysfunction Chronic hypoxemia and hypercapnia have been shown to cause increased circulating levels of norepinephrine, renin, and aldosterone and decreased levels of antidiuretic hormone. Renal arterial endothelium inCOPDpatients exhibits defects similar to those seen in the pulmonary arteries, shifting renal blood flow from the cortex to the medulla and impairing renal functional reserve. The combination of hemodynamic and hormonal disturbances leads to defective excretion of salt and water loads and, together with right ventricular dysfunction, to the plethoric and cyanotic manifestations of some patients with COPD.

Cachexia Weight loss sometimes occurs in patients with advancedCOPD. A body-mass index (BMI) < 25 kg/m_2 is associated with increased frequency of exacerbations and with significantly reduced survival. Cachexia has been attributed to caloric intake failing to keep pace with energy expenditures associated with increased work of breathing, but more recent evidence suggests that a biochemical basis is more likely. Hypoxemia leads to increased circulating levels of tumor necrosis factora (TNF-a), and weight loss has now been correlated with levels of the latter.

Peripheral Muscle Dysfunction Protein and muscle are lost as part of wasting in advancedCOPD. Skeletal muscle bulk is lost with proportional reductions in strength. Proximal limb girdle muscles of the upper and lower extremities are particularly affected, contributing to dyspnea with activities of daily living. Fiber composition in skeletal muscle changes, favoring endurance over strength. These changes occur in parallel with$FEV_1$and independently of glucocorticoid use, which can also cause myopathy and muscle weakness.

Osteoporosis Loss of bone density is common in advanced disease. Over half ofCOPDpatients lose more than 1 SD of bony density, and more than one-third have values more than 2 SDs below normal. Vertebral fractures are especially common. These changes are even more severe in patients receiving chronic glucocorticoid therapy.

NATURAL HISTORY

COPDis identified by the presence of an abnormalFEV_1in middle age, usually early in the fifth decade, and is characterized by an accelerated decline of $FEV_1$with aging. In normal individuals, FEV_1normally reaches a lifetime peak at age 25 and undergoes a linear decline of about 35 mL per year thereafter. Annual loss of FEV_1among susceptible individuals who develop COPD is between 50 and 100 mL per year. Greater rates of decline have been associated with mucus hypersecretion, especially in men, and with bronchial hyperreactivity. Acute exacerbations do not alter the rate of decline. Dyspnea and impairment of physical work capacity are characteristic only of moderately severe to severe airways obstruction. There is considerable variation among individual patients. The majority of patients usually experience exertional dyspnea when FEV_1falls

below 40% of predicted and have dyspnea at rest when the FEV_1<25% of predicted. In addition to dyspnea at rest, CO_2retention and cor pulmonale frequently occur when the FEV_1falls to 25% of predicted. With a respiratory infection, small changes in the degree of obstruction can make a large difference in symptoms and gas exchange. Thus small therapeutic gains may have rewarding results.

Exacerbation The clinical course ofCOPD can be characterized as one of slow progression and relative stability punctuated by episodic exacerbations occurring, on average, a little more than once per year. Exacerbations are generally described as a worsening of previously stable disease characterized by increased dyspnea, wheeze, and cough and sputum volume, tenacity, and purulence, with variable degrees of water retention and with worsening gas exchange and ventilation-perfusion relationships. Hyperinflation and work of breathing are increased. To the extent that diaphragmatic function and neuromuscular drive can compensate for the increased work, Pa_{CO_2}will not rise, but when work demands exceed respiratory pump capacity, hypercapnia and respiratory acidemia ensue. Cardiac output often does not increase sufficiently to compensate for the increased oxygen consumption from respiratory muscles, thereby compounding the hypoxemia due to / mismatching and hypercapnia.

MostCOPDexacerbations are thought to be a consequence of acute tracheobronchitis, usually infectious. Most infections are primarily bacterial or the consequence of bacterial superinfection of a primary viral process. Exacerbations may also be triggered by, and must be distinguished from, left ventricular failure, cardiac arrhythmias, pneumothorax, pneumonia, and pulmonary thromboembolism. Upper airway obstruction, aspiration, rhinitis or sinusitis, asthma, or gastroesophageal reflux should be excluded. Although COPD exacerbations are individually serious and potentially life-threatening, they do not cause accelerated declines ofFEV_1over time.

CLINICAL MANIFESTATIONS

HISTORY

Patients withCOPD are most often tobacco smokers with a history of at least one pack per day for at least 20 years. The disease is only rarely seen in nonsmokers. Onset is typically in the fifth decade and often comes to attention as a productive cough or acute chest illness. Exertional dyspnea is usually not encountered until the sixth or seventh decade. The patient's perception of dyspnea correlates poorly with physiologic measurements, especially among older patients. A morning "smoker's cough" is frequent, usually mucoid in character but becoming purulent during exacerbations, which in early disease are intermittent and infrequent. Volume is generally small. Production of more than 60 mL/d should prompt investigation for bronchiectasis. The frequency and severity of cough generally do not correlate with the degree of functional impairment. Wheezing may be present but does not indicate severity of illness. As COPD progresses, exacerbations become more severe and more frequent. Gas exchange disturbances, worsen and dyspnea becomes progressive. Exercise tolerance becomes progressively limited. With worsening hypoxemia, erythrocytosis and cyanosis may occur. The development of morning headache may indicate the onset of significant CO_2retention. In advanced disease, weight loss is frequent and correlates with an adverse prognosis. When blood gas derangements are severe, cor pulmonale may

manifest itself by peripheral edema and water retention. Anxiety, depression, and sleep disturbances are not infrequent.

PHYSICAL FINDINGS

The physical examination has poor sensitivity and variable reproducibility inCOPD. Findings may be minimal or even normal in mild disease, requiring objective laboratory data for confirmation. In early disease, the only abnormal findings may be wheezes on forced expiration and a forced expiratory time prolonged beyond 6 s. With progressive disease, findings of hyperinflation become more apparent. These include an increased anteroposterior diameter of the chest, inspiratory retraction of the lower rib margins (Hoover's sign), decreased cardiac dullness, and distant heart and breath sounds. Coarse inspiratory crackles and rhonchi may be heard, especially at the bases. To gain better mechanical advantage for their compromised respiratory muscles, patients with severe airflow obstruction may adopt a characteristic tripod sitting posture with the neck angled forward and the upper torso supported on the elbows and arms. Breathing through pursed lips prolongs expiratory time and may help reduce dynamic hyperinflation.

Cor pulmonale and right heart failure may be evidenced by dependent edema and an enlarged, tender liver (Chap. 237). With pulmonary hypertension, a loud pulmonic component of the second heart sound may be audible, along with a right ventricular heave and a murmur of tricuspid regurgitation; these findings may be obscured by hyperinflation. If right-sided pressures are sufficiently high, neck veins may elevate instead of collapse with inspiration (Kussmaul's sign). Cyanosis is a somewhat unreliable manifestation of severe hypoxemia and is seen when severe hypoxemia and erythrocytosis are present.

Radiographic Findings A posteroanterior and lateral chest film should be obtained primarily to exclude competing diagnoses. They may be entirely normal in mild disease. AsCOPDprogresses, abnormalities reflect emphysema, hyperinflation, and pulmonary hypertension. Emphysema is manifested by an increased lucency of the lungs. In smokers, these changes are more prominent in the upper lobes, while ina_1ATdeficiency, they are more likely in basal zones. Local radiolucencies>1 cm in diameter and surrounded by hairline arcuate shadows indicate the presence of bullae and are highly specific for emphysema. With hyperinflation, the chest becomes vertically elongated with low flattened diaphragms. The heart shadow is also vertical and narrow. The retrosternal airspace is increased on the lateral view, and the sternal-diaphragmatic angle exceeds 90°. In the presence of pulmonary hypertension, the pulmonary arteries become enlarged and taper rapidly. The right heart border may become prominent and impinge on the retrosternal airspace. The presence of "dirty lung fields" may reflect the presence of bronchiolitis.

Computed tomography has greater sensitivity and specificity for emphysema than the plain film but is rarely necessary except for the diagnosis of bronchiectasis and evaluation of bullous disease. Nonhomogeneous distribution of emphysema is thought by some to be an indicator of suitability for lung volume reduction surgery (LVRS).

PULMONARY FUNCTION TESTING (See alsoChap. 250)

Because of the imprecision of clinical findings, objective evaluation of the presence, severity, and reversibility of airflow obstruction is essential in the diagnostic evaluation ofCOPD. A normalFEV$_1$essentially excludes the diagnosis. The spirogram in COPD shows decreased volume changes with time and a failure to reach a plateau after 3 to 5 s. Continued airflow may be evident for 10 s or more on forced exhalation. The flow-volume curve shows diminished expiratory flow at all lung volumes. Expiratory flow is concave to the volume axis. When flow is plotted against absolute lung volume, the entire curve is shifted to higher volumes, reflecting hyperinflation. Serial spirometry is important in assessing the rate of decline of FEV$_1$.

Reversibility is assessed by spirometry before and after administration of an inhaled bronchodilator, most often a short-actingb$_2$-adrenergic agonist. Testing should be performed when the patient is clinically stable. Short-acting bronchodilators should be withheld for 6 h, long-acting dilators for 12 h, and theophylline for 24 h prior to testing. A significant response is an increase of at least 12% and 200 mL in eitherFEV$_1$or forced vital capacity (FVC). Postbronchodilator FEV$_1$ is useful for prognostication. Although only one-third ofCOPDpatients show a significant response to an inhaled bronchodilator in the pulmonary function laboratory on any one day, two-thirds will show a significant response when tested with different bronchodilators on several different occasions. The degree of bronchodilator response at any one testing session does not predict the degree of clinical benefit to the patient. Therefore, bronchodilators are given irrespective of the acute response obtained in the pulmonary function laboratory. The American Thoracic Society recommends staging COPD by FEV$_1$. Stage I, mild disease, is defined as FEV$_1$[3] 50% predicted; stage II, moderate disease, 35 to 49% predicted; and stage III, severe disease,<35% predicted.

Lung volumes are useful for the assessment of hyperinflation. Transfer factor for carbon monoxide (DL$_{CO}$) correlates negatively with the degree of emphysema but is not specific and may miss mild disease. Neither test is indicated routinely, but DL$_{CO}$ may help distinguish chronic asthma from emphysema.

Measurements for arterial blood gas are not needed for mild disease, but they should be assessed routinely for stage II or stage IIICOPD. Patients with pulmonary hypertension or cor pulmonale with normal daytime blood gases should be evaluated for nocturnal desaturation by overnight oximetry. Polysomnography to exclude concurrent sleep apnea should be obtained for patients who also complain of excessive daytime somnolence or who have a history of snoring.

a$_1$ATlevels are not needed routinely but should be obtained for chronic airflow obstruction or chronic bronchitis in nonsmokers, as well as inCOPDpatients with bronchiectasis, cirrhosis without apparent risks, premature emphysema, or basilar emphysema; in patients under age 50 with unremitting asthma; and in individuals with a family history ofa$_1$AT deficiency.

TREATMENT

Treatment of COPDis based on the principles of prevention of further evolution of disease, preservation of airflow, preservation and enhancement of functional capacity,

management of physiologic complications, and avoidance of exacerbations.

Smoking Cessation (See also Chap. 390) The Lung Health Study has demonstrated that elimination of tobacco smoking confers significant survival benefit to patients with COPD. Prolonged survival is associated with reduced rates of malignancy and cardiovascular disease as well as with a significant increment in FEV_1 in the first year after smoking cessation. The rate of decline of FEV_1 reverts back to that of a nonsmoker. Although bronchodilator therapy produces similar first-year gains in FEV_1, pharmacotherapy alone does not modify the decline of airflow over time. Even unsuccessful quitters show significant benefits when compared to continuing smokers.

Despite the demonstrated benefits of smoking cessation, sustained quitting is difficult to achieve. Overall, only 6% of smokers succeed in quitting long term, and 70 to 80% of short-term quitters start smoking again. Successful quitting requires concerted active and continuing intervention by the physician. The physician should address the issue in regular patient visits, assess the patient's readiness to quit, advise the patient as to the best methods for smoking cessation, provide emotional and pharmacologic support, and arrange close follow-up of the patient's efforts. The concept of "lung age" may be helpful in promoting smoking cessation by determining the age at which the observed FEV_1 would be a normal finding. Lungs of 50- to 60-year-old smokers may be "normal" for a 70- to 80-year-old individual. Nicotine patches and nicotine polacrilex gum improve quit rates, especially among nicotine-dependent smokers. The addition of oral bupropion at 150 mg twice daily produces significant additional benefit, with a 1-year sustained abstinence rate of 22.5% compared to 6% for placebo. Smoking cessation is typically associated with weight gain of 3 to 4 kg. To minimize weight gain, reluctance to quit, and relapse, prospective quitters should be counseled to reduce caloric intake and to increase physical activity.

Bronchodilators These drugs improve dyspnea and exercise tolerance by improving airflow and by reducing end-expiratory lung volume and air-trapping. Although airflow limitation is relatively fixed, some degree of response to bronchodilator medication is usually present. Bronchodilator medication is available in metered-dose inhaler (and some dry-powder inhalers) and in nebulizable and oral forms. Inhalers deliver medications directly to the airways and have limited systemic absorption and side effects. Proper use requires timing and coordination of inspiration and inhaler actuation and presents frequent difficulties for chronic lung patients. These problems can usually be overcome with education and with the use of holding chambers. Aerosol nebulizers have no pharmacologic advantage over metered-dose inhalers. Their use should be limited to patients who remain unable to master metered dose inhalers adequately. Oral medication is associated with higher rates of adherence than inhalers but shows higher rates of systemic side effects without superior bronchodilation.

Three major classes of bronchodilators are commonly employed in the treatment of patients with COPD: short- and long-acting β_2-adrenergic agonists, anticholinergics, and theophylline derivatives. Short-acting β_2-agonists (albuterol, pirbuterol, terbutaline, metaproterenol) are relatively bronchoselective with minimal effects on heart rate and blood pressure. They produce significant bronchodilation at 5 to 15 min and remain effective for 4 to 6 h. Long-acting β_2-agonists (oral sustained-release albuterol and inhaled salmeterol) have an onset of action of 15 to 30 min and a 12-h duration of

action. Anticholinergic agents (ipratropium bromide) have a 30- to 60-min onset of action and a 4- to 6-h duration. Theophyllines are generally administered orally in 12- or 24-h preparations. Recommended bronchodilator regimens are shown in Table 258-1.

Regular use of ipratropium may lead to improvements in baseline FEV_1 when compared with short-acting β_2-agonists. When used together, ipratropium and short-acting β_2-agents show greater clinical efficacy than either agent alone, without an increase in side effects. Salmeterol as a single agent produces longer lasting bronchodilation than ipratropium, improves baseline FEV_1 over time, and is not associated with loss of efficacy over a period of several months. Salmeterol, however, has not yet been evaluated as a component of combination therapy.

Theophylline is a weak bronchodilator with a narrow therapeutic window. Much of its clinical benefit derives from effects other than bronchodilation; therapeutic doses of theophylline increase ventilatory drive, enhance diaphragmatic contractility, and increase cardiac output. About 20% of COPD patients respond to theophylline with improved airflow, exercise tolerance, and quality of life. Theophylline produces additional benefits in exercise capacity and quality of life when used in combination with short-acting β_2-adrenergic agonists. The therapeutic range for theophylline is commonly given as 10 to 20 ug/mL, with greater efficacy but greater toxicity seen at higher serum levels. The risk of toxicity is greater in older patients and in those with heart and kidney disease. Optimal dosing must balance the competing considerations of risk and benefit for each individual patient.

Glucocorticoids Because COPD, like asthma, is a disease associated with airway inflammation, glucocorticoids are an intuitively attractive therapeutic modality. Nevertheless, results of clinical trials of glucocorticoid therapy in COPD patients have shown less impressive benefits when compared to patients with asthma. The degree of response to glucocorticoids appears to correlate with the presence of asthmatic features, but data supporting their use is limited. Only 10% more patients show subjective benefit and increase their FEV_1 or forced vital capacity by at least 20% when compared to those on placebo. Responders cannot be reliably identified on clinical grounds, although response to an inhaled β_2-agonist is commonly used as a predictor. The benefits of a 10- to 14-day trial of 30 to 40 mg/d of prednisone for patients with stage III disease who have not responded adequately to mixed bronchodilator therapy remain to be proven. Long-term systemic glucocorticoid use is associated with multiple side effects. In particular, they have been associated with worsened osteoporosis and increased risk of vertebral fracture. If systemic steroids are used, the lowest effective dose should be employed and alternate-day dosing used whenever possible. The use of inhaled glucocorticoids ameliorates systemic side effects. Three large clinical trials have shown that inhaled glucocorticoids do not alter the rate of decline of FEV_1. While an inhaled glucocorticoid does not decrease the number or frequency of COPD exacerbations, it may decrease their severity and reduce the need for hospitalization. Symptoms and exercise tolerance improve on inhaled glucocorticoids.

Management of α_1AT Deficiency Given the central role of smoking in the pathogenesis of disease, smoking cessation is an important cornerstone in the management of α_1AT deficiency. Exogenous α_1AT derived from pooled human plasma administered intravenously in a weekly dose of 60 mg/kg has been shown to induce protective levels

ofa1AT in deficient individuals. Because of the expense and inconvenience of the treatment, replacement ofa1AT is used only for patients over age 18 witha1AT levels below 11 umol/L who have stopped smoking and who have airflow obstruction. A recently published large nonrandomized trial showed that augmentation therapy significantly decreased 5-year mortality (RR 0.64) for patients receiving replacement. The rate of decline ofFEV1also decreased with augmentation therapy. In both instances, benefit was largely restricted to those patients with FEV1 35 to 49% of predicted. These findings require confirmation in randomized controlled trials.

Oxygen Severe and progressive hypoxemia is often seen in advancedCOPD and may result in cellular hypoxia with deleterious physiologic consequences. The establishment of adequate systemic oxygen transport is essential to the prevention of tissue hypoxia and requires attention to cardiac output and hemoglobin concentration as well as to arterial O2saturation (SaO2). Long-term O2therapy has been shown to reverse secondary polycythemia; improve body weight; ameliorate cor pulmonale; and enhance neuropsychiatric function, exercise tolerance, and activities of daily living. Two major studies, one in the United States and one in the United Kingdom, established a survival benefit for long-term O2therapy that increased with the number of hours per day that O2 was used. The mechanism for this benefit has not been conclusively elucidated, but it appears to be related to the stabilization of pulmonary hemodynamics.

The need for long-term O2therapy should be documented with measurement of arterial blood gases obtained at rest and confirmed by a separate determination of resting arterial blood gases during a period of medical stability after 30 to 90 days of optimum medical therapy. Once the need for O2 has been demonstrated in a stable patient, the requirement is generally for the duration of the patient's life. Patients with a PaO2£ 55 mmHg or SaO2£ 88% should be provided with oxygen titrated to raise SaO2 to³90%. Oxygen is likewise indicated for patients who have a PaO2 of 56 to 59 mmHg with SaO2³ 89% when hematocrit is >55% or when cor pulmonale or other objective evidence of pulmonary hypertension is present. Oxygen may be appropriate for patients whose resting awake PaO2³60 mmHg with SaO2³ 90% if they become hypoxic during exercise or sleep. Once oxygen is prescribed, the dose should be titrated to maintain SaO2³ 90% during sleep and normal walking, as well as at rest, and it should be used for a minimum of 15 h a day to realize a survival benefit.

Oxygen is most frequently delivered through a nasal cannula at rates of 2 to 5 L/min. Oxygen-sparing cannulae are available. Transtracheal administration provides further O2-sparing benefits but requires scrupulous attention to catheter maintenance and hygiene and is not suitable for all patients. Oxygen is packaged as compressed gas or compressed liquid or can be delivered from an O2concentrator, a molecular sieve that enriches O2 by removing nitrogen from ambient air. O2should be prescribed from sources that are appropriate to the individual patient's life-style and needs. It is customary to provide a stationary O2source, either an O2concentrator, which is dependent on a reliable source of electricity, or 100-kg (200-lb) H cylinders of compressed O2. Flow resistance imposes a 15-m (50-ft) practical limit to the length of tubing connecting the O2source to the patient's cannula. For patients whose activities of daily living require ambulation beyond this limit, ambulatory or portable systems should be provided. Ambulatory O2 needs may be met with rolling 10-kg (22-lb) E cylinders of compressed O2, or with portable 2-kg (4.5-lb) aluminum cylinders or 3-kg (6.6-lb) liquid

oxygen packs. The duration of O_2availability from an O_2concentrator is unlimited. For compressed gas and liquid sources, the amount of available oxygen is determined by the size of the system and the patient's liter flow needs. Portable systems generally provide 4 to 5 h of O_2flow.

Oxygen therapy is generally safe. Cylinders should be secured to prevent tipping over or potentially explosive disconnection of the regulator valve. Oxygen should be stored away from open flames or other source of heat, and patients and family members should be educated to be especially scrupulous about avoiding smoking in the presence of flowing O_2.

Prophylaxis No evidence supports the prophylactic use of antibiotics in stableCOPD. Yearly influenza vaccination is recommended for all patients with chronic cardiopulmonary disease, although objective benefit has not been conclusively demonstrated. Pneumococcal vaccination with 23-valent polysaccharide is also recommended. Amantadine should be used for unvaccinated patients who are placed at risk by an outbreak of influenza A.

Rehabilitation Airflow limitation, dyspnea, and muscle loss and deconditioning all compromise cardiopulmonary fitness and contribute to a progressively constrained daily life and unsatisfactory quality of life. Pulmonary rehabilitation is a multidisciplinary program of care for patients with chronic respiratory impairments that is individually tailored and designed to optimize physical and social performance. A pulmonary rehabilitation program consists of exercise training, patient education, psychosocial and behavioral intervention, and regular assessment of outcomes and is designed to minimize the disability and handicap imposed by the physiologic impairments consequent toCOPD. Rehabilitation in COPD should be considered for patients with persistent symptoms and disability despite optimal medical management. Spirometric criteria should not be the primary basis for referral into rehabilitation programs. Exercise consists of 20 to 30 min of upper and lower extremity exercise at 60 to 75% maximumO_2 or heart rate two to five times a week. Both strength and endurance exercises are provided. Education covers pursed lip and other breathing strategies to minimize dyspnea, energy-conservation skills, principles of medications and proper use of metered-dose inhalers, nutrition, and end-of-life decision-making. Behavioral interventions focus on dyspnea, depression, and self-sufficiency and on issues of control, coping, and role function. Dyspnea, exercise tolerance, activity level, and quality of life are followed at regular intervals. Pulmonary rehabilitation programs have been shown to improve endurance time for submaximal exercise by 38 to 80% and 6-min walking distance by 80 to 113 m. Clinically meaningful reduction in dyspnea and improvement of quality of life have been reported. No clinical trials have been adequately designed to address the issue of survival benefit. Reductions in costs of care and resource consumption have not reached statistical significance.

Despite maximal medical therapy, whenCOPDprogresses to stage III and is complicated by hypercapnia or pulmonary hypertension, surgical approaches to treatment may be considered.

Transplantation (See alsoChap. 267) Owing to its frequency in the general population, emphysema is the most common indication for lung transplantation. Transplantation

should be actively considered for end-stage COPD patients when the prognosis from the disease is worse than the survival statistics for the surgery. Lung transplantation should be considered for COPD patients who, despite maximal medical therapy, have an FEV_1 < 25% predicted and with pulmonary hypertension or cor pulmonale. Precedence is given to those patients with a Pa_{CO2} of 55 mmHg and progressive deterioration. Asthma and other reversible airflow limitation must be excluded. Rehabilitation and long-term O_2 therapy, where appropriate, should be provided prior to transplant evaluation.

Lung Volume Reduction Surgery LVRS, or pneumectomy, is designed to relieve dyspnea and improve exercise function in severely disabled patients with stage III emphysema. At operation, severely emphysematous lung tissue is resected, leading to improvement in elastic recoil in the remaining pulmonary parenchyma. This decreases hyperinflation and enhances diaphragmatic function, with consequent 25 to 50% improvement of airflow and exercise capacity. In early uncontrolled studies, hospital mortality for LVRS ranged from 5 to 18% and hospital stays averaged 9 to 18 days, with frequent significant air leaks. Cost of LVRS was $33,000 to $70,000 per case. Because of the large number of potential candidates, the high cost involved, and unanswered questions about the benefits of the operation, use of LVRS in the United States has been restricted to a multicenter randomized controlled trial, the National Emphysema Treatment Trial (NETT), comparing LVRS with best medical therapy. Stage III emphysema patients accepted for evaluation into NETT are under age 75, are severely hyperinflated, and have severe dyspnea despite optimal medical therapy. Contraindications to LVRS are similar to those for lung transplantation, including active smoking, marked obesity or cachexia, and inability to undertake pulmonary rehabilitation successfully. There has been little consensus regarding features identifying ideal and suboptimal candidates for the surgery. Radiographic heterogeneity of disease and the absence of significant intrinsic airway disease have been suggested characteristics of patients likely to benefit. Results from the NETT suggest that physiologic benefits from LVRS may begin to be lost as early as 1 year after surgery. Accelerated declines of FEV_1 have been reported, averaging 100 mL per year and particularly marked in those patients with the greatest postoperative gains in airflow. Improvements in dyspnea and exercise tolerance may be sustained for as long as 3 years but may decline thereafter. Until these issues are satisfactorily resolved, LVRS will remain an experimental procedure.

Treatment of Exacerbations

Triage The initial decision in the management of an exacerbation of COPD is whether hospitalization is necessary. Rapidity of evolution of symptoms and response to initial therapy, level of consciousness, presence or absence of respiratory distress, severity of gas exchange disturbance, and arterial blood gas deviation from the patient's stable baseline should influence the decision to hospitalize. The patient's ability to manage at home and the resources available for home care should weigh heavily in the decision-making process.

Home Therapy For patients with mild exacerbations for whom outpatient therapy is appropriate, a combination of anticholinergic and short-acting b_2-adrenergic agonist bronchodilators should be prescribed. Although b_2-agonists may be given as frequently as once an hour, there is no advantage to administering anticholinergic bronchodilators

more frequently than every 4 to 6 h. Metered-dose inhalers should be used with spacers. There is no evidence that the use of nebulizers provides any improvement in outcome.

The presence of increased sputum volume or purulence suggest an infectious cause of an exacerbation. With either of these features is present in conjunction with increased breathlessness or when both are present, antibiotics should be prescribed. The organisms most frequently associated with mildCOPDexacerbations include *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*. Trimethoprim/sulfamethoxazole, doxycycline, or amoxicillin is an appropriate management option, although choices may be modified by local antibiotic sensitivity data.

There is a need for well-controlled studies on the utility of glucocorticoids in the outpatient management ofCOPDexacerbations. Oral glucocorticoids may be continued in patients already receiving such treatment or given to patients who do not show a satisfactory response to bronchodilator therapy. The usual dose is 20 to 40 mg daily for 7 to 10 days. Short-term glucocorticoid therapy lasting less than 3 weeks may be discontinued without the use of a tapering dose.

Hospital Management For patients with exacerbations of sufficient severity to warrant hospitalization, improvement of airflow, gas exchange, and acid-base status are of central importance. Hospitalized patients should receive bronchodilators, antibiotics, oral glucocorticoids, and sufficient O_2 to keep the Sa_{O2}³ 90%.b_2-agonists and anticholinergic agents should be given together every 4 to 6 h. The frequency of sympathomimetic bronchodilator administration may be increased as needed to as often as every 20 min. Because high doses ofb_2-agonists may cause hypokalemia, serum potassium levels and heart rate should be monitored closely for patients receiving frequent doses of these agents. Data are contradictory regarding the addition of theophylline to the bronchodilator regiment of patients showing an inadequate initial response, yet the current American and British Thoracic Societies' guidelines recommend consideration of its use to produce plasma theophylline levels between 10 and 20 ug/mL. Oral glucocorticoids have been shown to produce modest improvements in$FEV_1$and in the duration of hospitalization forCOPDexacerbations. Recent data indicate that more severe COPD exacerbations are associated with the recovery of enterobacteriaciae in respiratory secretions. For this reason, a second- or third-generation cephalosporin, a fluoroquinolone, a second-generation macrolide, or an extended-spectrum penicillin is now recommended as initial therapy. Attempts to obtain diagnostically adequate sputum should be made, and, when available, sputum results should be used to individualize therapy in the light of local microbial sensitivity spectra. Oxygen therapy is an important component of the management of a severe exacerbation of COPD. It is important to maintain the Sa_{O2}> 90% and Pa_{O2}between 60 and 65 mmHg for most patients. In many cases, administration of O_2 will result in worsening hypercapnia, although rarely to a clinically significant degree if the O_2 is used only in amounts to achieve the minimal goals. The elevation of Pa_{CO2} is multifactorial, resulting from increased dead space due to reduced tidal volume as well as from the Haldane effect, i.e., a right wave shift of the CO_2dissociation curve in the presence of increased saturated hemoglobin. The lower the initial Pa_{O2} and the greater the increase, the larger the increase in Pa_{CO2}observed. Patients whose pH on presentation is below

7.25 and with Pa_{O_2}< 50 mmHg are at particular risk and should be observed closely.

For patients at increased risk of hypercapnia, administration of controlled concentrations of O_2through a Venturi mask is reasonable. Inspired O_2concentrations (FI_{O_2}) of 0.24 to 0.28 are usually sufficient to keep Sa_{O_2}³90%.

*MECHANICAL VENTILATION (See also*Chap. 266*)* Patients with impaired consciousness, respiratory distress evidenced by tachypnea with a respiratory rate greater than 35 breaths per minute and/or abdominal paradox, severe hypoxemia, or significant respiratory acidosis with pH< 7.25 and who deteriorate despite treatment are candidates for immediate ventilatory support using either noninvasive (mask) or invasive (intubation) approaches. The goals are to buy time for medical treatments to take effect, to rest the respiratory muscles, and to improve gas exchange abnormalities while avoiding the major complications of mechanical ventilatory support.

Noninvasive positive-pressure ventilation (NIPPV) delivered by nasal mask should be considered in units that have experience with the technique for patients who remain alert and cooperative, who are not heavily sedated, who are hemodynamically stable, and who are able to clear their airways by coughing up secretions. In these circumstances, NIPPV has been shown to be successful in avoiding the need for endotracheal intubation in up to 70% of cases. Success, as evidenced by improved $Pa_{CO_2}$and pH, should be evident within the first 60 min. Part-time NIPPV for 6 to 8 h per day may afford sufficient respiratory muscle rest to avert the need for invasive conventional ventilation. Failed attempts at NIPPV can be followed by intubation and conventional ventilation and do not appear to carry a worse prognosis. Successful application of NIPPV has been associated with a decrease in intensive care and hospital stays, incidence of nosocomial pneumonia, and costs.

Before committing to endotracheal intubation and conventional ventilatory support, the patient's wishes for such support, the patient's quality of life, and the benefits and costs of care should be thoroughly reviewed. Where the patient's wishes cannot be clearly ascertained or there is uncertainty about the appropriateness of the intervention, intubation and ventilation should proceed. If mechanical ventilatory support is subsequently determined to be inappropriate, support may then be withdrawn.

Once intubation is accomplished, the patient can be ventilated in the controlled ventilation, assist-control, intermittent mandatory ventilation, or pressure support modes. FI_{O_2}should be sufficient to obtain Sa_{O_2}³ 90% and Pa_{O_2} of 60 to 65 mmHg. An FI_{O_2} of 0.24 to 0.40 is usually adequate for the purpose. Minute volume should be adequate to keep pH ³ 7.25, but one should not strive to achieve a "normal" Pa_{CO_2}. It is important to try to avoid overventilation and hyperinflation in ventilatedCOPDpatients. Because the time constant for exhalation is abnormally prolonged, it is essential to allow adequate expiratory time to permit as complete emptying of each breath as possible, preferably at least 3 to 4 s. This is best accomplished by minimizing tidal volume and respiratory rate. Lesser gains in expiratory (E) time can be obtained by high inspiratory (I) flow rates and I:E ratios of 1:2 or higher. Inadequate expiratory time leads to dynamic hyperinflation and in turn to the development of intrinsic positive end-expiratory pressure (PEEPi). PEEPi is just as capable of producing hypotension as extrinsically applied PEEP. When a mechanically ventilated patient with obstructive lung disease abruptly develops

hypotension, PEEPi should be excluded, either by direct measurement or by disconnecting the patient from the ventilator for 30 to 60 s. PEEPi and dynamic hyperinflation increase the work of breathing, place the diaphragm at mechanical disadvantage, and contribute significantly to difficulties in weaning from ventilatory support. Over a period of days, as the underlying precipitants of the exacerbation are controlled, airway obstruction gradually remits and gas exchange improves and it becomes appropriate to consider removal from mechanical ventilatory support.

The principles of weaning from mechanical ventilation are discussed in detail in Chap. 266.

Prognosis after Exacerbation The hospital mortality rate for an episode of respiratory failure in COPD ranges from 11 to 25% and depends on the severity of the episode, the patient's chronic health and nutritional status, and the presence of cor pulmonale or congestive heart failure. Data regarding subsequent course may be helpful in educating COPD patients and in guiding their subsequent management decisions. Among survivors of mechanical ventilation, the 6-month mortality rate is approximately 40%. Two-thirds of survivors have frequent recurrences of exacerbations, and functional status thereafter is often poor.

(Bibliography omitted in Palm version)

Back to Table of Contents

259. INTERSTITIAL LUNG DISEASES - *Talmadge E. King, Jr.*

The interstitial lung diseases (ILDs) represent a large number of conditions that involve the parenchyma of the lung -- the alveoli, the alveolar epithelium, the capillary endothelium, and the spaces between these structures, as well as the perivascular and lymphatic tissues. This heterogeneous group of disorders is classified together because of similar clinical, roentgenographic, physiologic, or pathologic manifestations. These disorders are often associated with considerable morbidity and mortality, and there is little consensus regarding the best management of most of them.

ILDshave been difficult to classify because more than 200 known individual diseases are characterized by diffuse parenchymal lung involvement, either as the primary condition or as a significant part of a multiorgan process, as may occur in the connective tissue diseases (CTDs). One useful approach to classification is to separate the ILDs into two groups, those of known and those of unknown causes (Table 259-1). Each of these groups can be subdivided into subgroups according to the presence or absence of histologic evidence of granulomas in interstitial or vascular areas. For each ILD there may be an acute phase, and there is usually a chronic one as well. Rarely, some are recurrent, with intervals of subclinical disease.

Sarcoidosis (Chap. 318), idiopathic pulmonary fibrosis (IPF), and pulmonary fibrosis associated withCTDs(Chaps. 311 to 317) are the most commonILDs of unknown etiology. Among the ILDs of known cause, the largest group comprises occupational and environmental exposures, especially the inhalation of inorganic dusts, organic dusts, and various fumes or gases (Chaps. 253 and254). A clinical diagnosis is possible for many forms of ILD, especially if an occupational and environmental history is aggressively pursued. For other forms, tissue examination, usually obtained by thoracoscopic or open lung biopsy is critical to confirmation of the diagnosis. High-resolution computed tomography (HRCT) scanning promises to improve diagnostic accuracy further as histologic-image correlation is perfected.

PATHOGENESIS

TheILDs are nonmalignant disorders and are not caused by identified infectious agents. The precise pathway(s) leading from injury to fibrosis is not known. Although there are multiple initiating agent(s) of injury, the immunopathogenic responses of lung tissue are limited, and the mechanisms of repair have common features. Two major histopathologic patterns are found in patients with ILD: a granulomatous pattern (Fig. 259-1) and a pattern in which inflammation and fibrosis predominate.

GRANULOMATOUS LUNG DISEASE

This process is characterized by an accumulation of T lymphocytes, macrophages, and epithelioid cells organized into discrete structures (granulomas) in the lung parenchyma. The granulomatous lesions can progress to fibrosis. Many patients with granulomatous lung disease remain free of severe impairment of lung function, or, when symptomatic, they improve after treatment. The main differential diagnosis is between sarcoidosis (Chap. 318) and hypersensitivity pneumonitis (Chap. 253).

INFLAMMATION AND FIBROSIS

The initial insult is an injury to the epithelial surface causing inflammation in the air spaces and alveolar walls. If the disease becomes chronic, inflammation spreads to adjacent portions of the interstitium and vasculature and eventually causes interstitial fibrosis. Other important histopathologic patterns inILDsinclude diffuse alveolar damage (acute or organizing), desquamative interstitial pneumonia, respiratory bronchiolitis, lymphocytic interstitial pneumonia, and an organizing pneumonia [bronchiolitis obliterans with organizing pneumonia (BOOP) pattern]. The development of irreversible scarring (fibrosis) of alveolar walls, airways, or vasculature is the most feared outcome in all of these conditions because it is often progressive and leads to significant derangement of ventilatory function and gas exchange.

INITIAL EVALUATION

Patients withILDscome to medical attention mainly because of the onset of progressive exertional dyspnea or a persistent, nonproductive cough. Hemoptysis, wheezing, and chest pain may be present. Often, the identification of interstitial opacities on chest x-ray focuses the diagnostic approach toward one of the ILDs.

HISTORY

Duration of Illness *Acute presentation* (days to weeks), while unusual, occurs with allergy (drugs, fungi, helminths), acute idiopathic interstitial pneumonia, eosinophilic pneumonia, and hypersensitivity pneumonitis. These conditions may be confused with atypical pneumonias because of diffuse alveolar opacities on chest x-ray. *Subacute presentation* (weeks to months) may occur in all ILDs but is seen especially in sarcoidosis, drug-induced ILDs, the alveolar hemorrhage syndromes, cryptogenic organizing pneumonia (COP), and the acute immunologic pneumonia that complicates systemic lupus erythematosus (SLE) or polymyositis. In most ILDs the symptoms and signs are *chronic* (months to years). Examples includeIPF, sarcoidosis, pulmonary Langerhans cell histiocytosis (PLCH) (also known as Langerhans cell granulomatosis, eosinophilic granuloma, and histiocytosis X), pneumoconioses, andCTDs. *Episodic presentations* are unusual and include eosinophilic pneumonia, hypersensitivity pneumonitis, cryptogenic organizing pneumonia, vasculitides, pulmonary hemorrhage, and Churg-Strauss syndrome.

Age Most patients with sarcoidosis,ILDassociated withCTD, lymphangioleiomyomatosis (LAM),PLCH, inherited forms of ILD (familialIPF, Gaucher's disease, Hermansky-Pudlak syndrome) present between the ages of 20 and 40 years. Most patients with IPF are older than 50 years.

Gender LAM and pulmonary involvement in tuberous sclerosis occur exclusively in premenopausal women. Also,ILD in Hermansky-Pudlak syndrome and in theCTDs is more common in women; an exception is ILD in rheumatoid arthritis, which is more common in men. Because of occupational exposures, pneumoconioses also occur more frequently in men.

Family History Family history is occasionally helpful because familial associations (with

an autosomal dominant pattern) have been identified in tuberous sclerosis and neurofibromatosis. An autosomal recessive pattern of inheritance occurs in Niemann-Pick disease, Gaucher's disease, and the Hermansky-Pudlak syndrome. Familial clustering has been increasingly identified in sarcoidosis and familial pulmonary fibrosis, a process similar toIPF.

Smoking History Patients withPLCH, desquamative interstitial pneumonia (DIP), Goodpasture's syndrome, and respiratory bronchiolitis are almost always current or former smokers. Two-thirds to 75% of patients withIPF have a history of smoking.

Occupation and Environmental History A strict chronological listing of the patient's lifelong employment must be sought, including specific duties and known exposures. In hypersensitivity pneumonitis (Fig. 259-1), respiratory symptoms, fever, chills, and an abnormal chest roentgenogram are often temporally related to a hobby (pigeon breeder's disease) or to the workplace (Farmer's lung) (Chap. 253). Symptoms may diminish or disappear after the patient leaves the site of exposure for several days; similarly, symptoms may reappear on returning to the exposure site.

Other Important Past History Parasitic infections may cause pulmonary eosinophilia, and therefore a travel history should be taken in patients with known or suspectedILD. History of risk factors for HIV infection should be elicited from all patients with ILD because several processes may occur at the time of initial presentation or during the clinical course, e.g., HIV infection,BOOP, acute interstitial pneumonia, lymphocytic interstitial pneumonitis, or diffuse alveolar hemorrhage.

RESPIRATORY SYMPTOMS AND SIGNS

Dyspnea is a common and prominent complaint in patients withILD, especially the idiopathic interstitial pneumonias, hypersensitivity pneumonitis,COP, sarcoidosis, eosinophilic pneumonias, andPLCH. Some patients, especially patients with sarcoidosis, silicosis, PLCH, hypersensitivity pneumonitis, lipoid pneumonia, or lymphangitis carcinomatosis may have extensive parenchymal lung disease on chest x-ray without significant dyspnea, especially early in the course of the illness. Wheezing is an uncommon manifestation of ILD but has been described in patients with chronic eosinophilic pneumonia, Churg-Strauss syndrome, respiratory bronchiolitis, and sarcoidosis. Clinically significant chest pain is uncommon in most ILDs. However, substernal discomfort is common in sarcoidosis. Sudden worsening of dyspnea, especially if associated with acute chest pain, may indicate a spontaneous pneumothorax, which occurs in PLCH, tuberous sclerosis,LAM, and neurofibromatosis. Frank hemoptysis and blood-streaked sputum are rarely presenting manifestations of ILD but can be seen in the diffuse alveolar hemorrhage syndromes (DAHs), LAM, tuberous sclerosis, and the granulomatous vasculitides. Fatigue and weight loss are common in all ILDs.

PHYSICAL EXAMINATION

The findings are usually not specific. Most commonly, physical examination reveals tachypnea, and bibasilar end-inspiratory dry crackles, which are common in most forms ofILDassociated with inflammation but are less likely to be heard in the granulomatous

lung diseases. Crackles may be present in the absence of radiographic abnormalities on the chest radiograph. Scattered late inspiratory high-pitched rhonchi -- so-called inspiratory squeaks -- are heard in patients with bronchiolitis. The cardiac examination is usually normal except in the mid or late stages of the disease when findings of pulmonary hypertension and cor pulmonale may become evident (Chap. 237). Cyanosis and clubbing of the digits occurs in some patients with advanced disease.

LABORATORY

Antinuclear antibodies, anti-immunoglobulin antibodies (rheumatoid factors), and circulating immune complexes are identified in some patients, even in the absence of a definedCTD. A raised LDH is a nonspecific finding common to ILDs. Elevation of the serum angiotensin-converting enzyme level is common in sarcoidosis. Serum precipitins confirm exposure when hypersensitivity pneumonitis is suspected, although they are not diagnostic of the process. Antineutrophil cytoplasmic or anti-basement membrane antibodies are useful if vasculitis is suspected. The electrocardiogram is usually normal unless pulmonary hypertension is present; then it demonstrates right-axis deviation or right ventricular hypertrophy. Echocardiography also reveals right ventricular dilatation and/or hypertrophy in the presence of pulmonary hypertension.

CHEST IMAGING STUDIES

Chest X-ray ILD may be first suspected on the basis of an abnormal chest radiograph, which most commonly reveals a bibasilar reticular pattern. A nodular or mixed pattern of alveolar filling and increased reticular markings may also be present (seeFig. 249-1). A subgroup of ILDs exhibit nodular opacities with a predilection for the upper lung zones [sarcoidosis,PLCH, chronic hypersensitivity pneumonitis, silicosis, berylliosis, rheumatoid arthritis (necrobiotic nodular form), ankylosing spondylitis]. The chest x-ray correlates poorly with the clinical or histopathologic stage of the disease. The radiographic finding of honeycombing correlates with pathologic findings of small cystic spaces and progressive fibrosis; when present, it portends a poor prognosis. In most cases, the chest radiograph is nonspecific and usually does not allow a specific diagnosis.

Computed Tomography HRCT is superior to the plain chest x-ray for early detection and confirmation of suspectedILD. Also, HRCT allows better assessment of the extent and distribution of disease, and it is especially useful in the investigation of patients with a normal chest radiograph. Coexisting disease is often best recognized on HRCT scanning, e.g., mediastinal adenopathy, carcinoma, or emphysema. In the appropriate clinical setting HRCT may be sufficiently characteristic to preclude the need for lung biopsy inIPF, sarcoidosis, hypersensitivity pneumonitis, asbestosis, lymphangitic carcinoma, andPLCH. When a lung biopsy is required, HRCT scanning is useful for determining the most appropriate area from which biopsy samples should be taken.

Radionuclide Scanning Gallium-67 lung scanning is of limited value in evaluating the inflammatory component ofILD. An accelerated clearance from the lung of soluble aerosolized hydrophilic radionuclides such as^{99m}Tc-diethylenetriamene pentaacetate (DTPA) is an index of pulmonary epithelial permeability that results from inflammation. This test may provide a means of assessing the activity of ILD. Normal^{99m}Tc-DTPA

clearance inIPFpredicts stable disease, while rapid clearance identifies patients at risk for deterioration.

PULMONARY FUNCTION TESTING

Spirometry and Lung Volumes Measurement of lung function is important in assessing the extent of pulmonary involvement in patients withILD. Most forms of ILD produce a restrictive defect with reduced total lung capacity (TLC), functional residual capacity, and residual volume (Chap. 250). Forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) are reduced, but these changes are related to the decreased TLC. The FEV_1/FVC ratio is usually normal or increased. Reductions in lung volumes increase as lung stiffness worsens with disease progression. A few disorders (uncommon in sarcoidosis and hypersensitivity pneumonitis, while common in tuberous sclerosis and LAM) produce interstitial opacities on chest x-ray and obstructive airflow limitation on lung function testing.

Diffusing Capacity A reduction in the diffusing capacity of the lung for carbon monoxide D_{LCO} is a common but nonspecific finding in mostILDs. This decrease is due, in part, to effacement of the alveolar capillary units but, more importantly, to mismatching of ventilation and perfusion (/). Lung regions with reduced compliance due to either fibrosis or cellular infiltration may be poorly ventilated but may still maintain adequate blood flow and / in these regions act like true venous admixture. The severity of the reduction in D_{LCO}does not correlate with disease stage.

Arterial Blood Gas The resting arterial blood gas may be normal or reveal hypoxemia (secondary to a mismatching of ventilation to perfusion) and respiratory alkalosis. A normal arterial O_2tension (or saturation by oximetry) at rest does not rule out significant hypoxemia during exercise or sleep. CO_2retention is rare and is usually a manifestation of end-stage disease.

Cardiopulmonary Exercise Testing Because hypoxemia at rest is not always present and because severe exercise-induced hypoxemia may go undetected, it is useful to perform exercise testing with measurement of arterial blood gases to detect abnormalities of gas exchange. Arterial oxygen desaturation, a failure to decrease dead space appropriately with exercise [i.e., a high V_D/V_Tratio (Chap. 250)], and an excessive increase in respiratory rate with a lower-than-expected recruitment of tidal volume provide useful information about physiologic abnormalities and extent of disease. Serial assessment of resting and exercise gas exchange is an excellent method for following disease activity and responsiveness to treatment, especially in patients withIPF.

FIBEROPTIC BRONCHOSCOPY AND BRONCHOALVEOLAR LAVAGE (BAL)

In selected diseases (e.g., sarcoidosis, hypersensitivity pneumonitis,DAHs, cancer, pulmonary alveolar proteinosis), cellular analysis of BAL fluid may be useful in narrowing the differential diagnostic possibilities among various types ofILD. The role for BAL in defining the stage of disease and assessment of disease progression or response to therapy remains poorly understood, and the usefulness of BAL in the clinical assessment and management remains to be established.

TISSUE AND CELLULAR EXAMINATION

Lung biopsy is the most effective method for confirming the diagnosis and assessing disease activity. The findings may identify a more treatable process than originally suspected, particularly chronic hypersensitivity pneumonitis,COP, respiratory bronchiolitis-associatedILD, or sarcoidosis. Biopsy should be obtained before initiation of treatment. A definitive diagnosis avoids confusion and anxiety later in the clinical course if the patient does not respond to therapy or suffers serious side effects from it.

Fiberoptic bronchoscopy with multiple transbronchial lung biopsies (4 to 8 biopsy samples) is often the initial procedure of choice, especially when sarcoidosis, lymphangitic carcinomatosis, eosinophilic pneumonia, Goodpasture's syndrome, or infection are suspected. If a specific diagnosis is not made by transbronchial biopsy, then surgical lung biopsy by video-assisted thoracic surgery or open thoracotomy is indicated. Adequate-sized biopsies from multiple sites, usually from two lobes, should be obtained. Relative contraindications to lung biopsy include serious cardiovascular disease, "honeycombing" and other roentgenographic evidence of diffuse end-stage disease, severe pulmonary dysfunction, or other major operative risks, especially in the elderly.

TREATMENT

Although the course ofILD is variable, progression is common and often insidious. All treatable possibilities should be carefully considered. Since therapy does not reverse fibrosis, the major goals of treatment are permanent removal of the offending agent when known and early identification and aggressive suppression of the acute and chronic inflammatory process, thereby reducing further lung damage.

Hypoxemia (PaO_2<55 mmHg) at rest and/or with exercise should be managed by supplemental oxygen. If cor pulmonale develops, diuretic therapy and phlebotomy may occasionally be required (Chap. 237).

Drug Therapy Glucocorticoids are the mainstay of therapy for suppression of the alveolitis present inILD, but the success rate is low. There have been no placebo-controlled trials of glucocorticoids in ILD, so there is no direct evidence that steroids improve survival in many of the diseases for which they are commonly used. Glucocorticoid therapy is recommended for symptomatic ILD patients with idiopathic interstitial pneumonias, eosinophilic pneumonias,COP,CTD, sarcoidosis, acute inorganic dust exposures, acute radiation pneumonitis,DAH, and drug-induced ILD. In organic dust disease, glucocorticoids are recommended for both the acute and chronic stages.

The optimal dose and proper length of therapy with glucocorticoids in the treatment of mostILDs is not known. A common starting dose is prednisone, 0.5 to 1 mg/kg in a once-daily oral dose (based on the patient's lean body weight). This dose is continued for 4 to 12 weeks, at which time the patient is reevaluated. If the patient is stable or improved, the dose is tapered to 0.25 to 0.5 mg/kg and is maintained at this level for an additional 4 to 12 weeks depending on the course. Rapid tapering or a shortened course of glucocorticoid treatment can result in recurrence. If the patient's condition

continues to decline while on glucocorticoids, a second agent (see below) is often added and the prednisone dose is lowered to or maintained at 0.25 mg/kg per day.

Cyclophosphamide and azathioprine (1 to 2 mg/kg lean body weight per day) with or without glucocorticoids, have been tried with variable success inIPF, vasculitis, and otherILDs. An objective response usually requires at least 8 to 12 weeks to occur. In situations in which these drugs have failed or could not be tolerated, other agents, including methotrexate, colchicine, penicillamine, and cyclosporine, have been tried. However, their role in the treatment of ILDs remains to be determined.

Many cases ofILD are chronic and irreversible despite the therapy discussed above, and lung transplantation may then be considered (Chap. 267).

INDIVIDUAL FORMS OF ILD

IDIOPATHIC PULMONARY FIBROSIS

Several risk factors appear to be associated with the development ofIPF, a commonILD of unknown etiology. These include cigarette smoking; exposure to antidepressants; a history of chronic aspiration secondary to gastroesophageal reflux; and exposures to metal dust, wood dust, and solvents. Numerous viruses have been implicated in the pathogenesis of IPF, but no clear evidence for a viral etiology has been confirmed. The most compelling evidence for participation of genetic factors is the description fo familial cases of pulmonary fibrosis, which is transmitted as an autosomal dominant trait with variable penetrance. An association has been reported between IPF and a, antitrypsin inhibition (Pi) alleles on chromosome 14.

Clinical Manifestations Exertional dyspnea, a nonproductive cough, and inspiratory crackles with or without digital clubbing may be present on physical examination. The chest roentgenogram andHRCTtypically show patchy, predominantly peripheral, subpleural, reticular opacities in the lower lung zones. There may also be a ground-glass opacity usually associated with traction bronchiectasis and bronchiolectasis or subpleural honeycombing. Pulmonary function tests often reveal a restrictive pattern, a reduced D_{LCO}, and arterial hypoxemia that is exaggerated or elicited by exercise.

Histologic Findings Confirmation of the presence of the usual interstitial pneumonia (UIP) pattern on histologic examination is essential to confirm this diagnosis (Fig. 259-2). Transbronchial biopsies are not helpful in making the diagnosis of UIP, and surgical biopsy is usually required. The histologic hallmark and chief diagnostic criterion of UIP is a heterogeneous appearance at low magnification with alternating areas of normal lung, interstitial inflammation, fibrosis, and honeycomb changes. The latter are composed of cystic fibrotic air spaces that are frequently lined by bronchiolar epithelium and filled with mucin. Smooth muscle hyperplasia is commonly present in areas of fibrosis and honeycomb change. Biopsies taken from patients during an accelerated phase of their illness may show a combination of UIP and diffuse alveolar damage. These histologic abnormalities affect the peripheral, subpleural parenchyma most severely. The interstitial inflammation is usually patchy and consists of a lymphoplasmacytic infiltrate in the alveolar septa, associated with hyperplasia of type 2

pneumocytes. The fibrotic zones are composed mainly of dense collagen, although scattered foci of proliferating fibroblasts are a consistent finding. The extent of fibroblastic proliferation is predictive of disease progression. A UIP-like pattern can also be seen withCTDs, pneumoconioses (e.g., asbestosis), radiation injury, certain drug-induced lung diseases (e.g., nitrofurantoin), and chronic aspiration. Also, a fibrotic pattern may be found in the chronic stage of several specific disorders such as sarcoidosis, chronic hypersensitivity pneumonitis, organized chronic eosinophilic pneumonia, andPLCH. Since other histopathologic features are frequently present in these syndromes, the term UIP is used for those patients in whom the lesion is idiopathic and not associated with another condition.

TREATMENT

The clinical course is variable with a 5-year survival rate of 30 to 50% after diagnosis. Treatment options include glucocorticoids, cytotoxic agents (e.g., azathioprine, cyclophosphamide), and antifibrotic agents (e.g., colchicine, perfenidone, or interferon gamma-1b), alone or in combination with glucocorticoids. However, there is no firm evidence that any of these treatment approaches improves survival or the quality of life. Because of the poor prognosis in untreated patients, a therapeutic trial may be tried. If therapy is recommended, it should be started at the first identification of clinical or physiologic evidence of impairment of lung function. Lung transplantation should be considered for those patients who experience progressive deterioration despite optimal medical management and who meet the established criteria (Chap. 267).

DESQUAMATIVE INTERSTITIAL PNEUMONIA

DIPis a rare but distinct clinical and pathologic entity found exclusively in cigarette smokers. The histologic hallmark is the extensive accumulation of macrophages in intraalveolar spaces with minimal interstitial fibrosis. The peak incidence is in the fourth and fifth decades. Most patients present with dyspnea. Lung function testing shows a restrictive pattern with reduced DL_{CO} and arterial hypoxemia. The chest x-ray usually shows diffuse hazy opacities. Clinical recognition of DIP is important because the process is associated with a better prognosis (10-year survival rate is ~70%) and a better response to smoking cessation and systemic glucocorticoids than the more commonIPF. Respiratory bronchiolitis-associatedILD is considered to be a subset of DIP and is characterized by the accumulation of macrophages in peribronchical alveoli.

ACUTE INTERSTITIAL PNEUMONIA (AIP) (HAMMAN-RICH SYNDROME)

This is a rare, fulminant form of lung injury characterized by diffuse alveolar damage on lung biopsy. Most patients are older than 40 years. AIP is similar in presentation to the acute respiratory distress syndrome (ARDS) (Chap. 265) and probably corresponds to the subset of cases of idiopathic ARDS. The onset is usually abrupt in a previously healthy individual. A prodromal illness, usually lasting 7 to 14 days before presentation, is common. Fever, cough, and dyspnea are frequent manifestations at presentation. Diffuse, bilateral, air-space opacification is present on chest radiograph. HRCT scans show bilateral, patchy, symmetric areas of ground-glass attenuation. Bilateral areas of air-space consolidation may also be present. A predominantly subpleural distribution may be seen. The diagnosis of AIP requires the presence of a clinical syndrome of

idiopathic ARDS and pathologic confirmation of organizing diffuse alveolar damage. Therefore, lung biopsy is required to confirm the diagnosis. Most patients have moderate to severe hypoxemia and develop respiratory failure. Mechanical ventilation is often required. The mortality rate is high (>60%), with most patients dying within 6 months of presentation. Recurrences have been reported. However, those who recover often have substantial improvement in lung function. The main treatment is supportive. It is not clear that glucocorticoid therapy is effective.

NONSPECIFIC INTERSTITIAL PNEUMONIA (NSIP)

This condition defines a subgroup of the idiopathic interstitial pneumonias that can be distinguished clinically and pathologically fromUIP,DIP,AIP, and idiopathicBOOP. Lung biopsy shows varying proportions of chronic interstitial inflammation and fibrosis. NSIP is a subacute restrictive process that usually occurs at a younger age than UIP. It is often associated with a febrile illness, relative lack of clubbing, andHRCTfindings that show ground-glass opacities and areas of consolidation. Unlike patients withIPF, most patients with NSIP have a good prognosis, and most show improvement after treatment with glucocorticoids.

ILDASSOCIATED WITH CONNECTIVE TISSUE DISORDERS

Clinical findings suggestive of aCTD(musculoskeletal pain, weakness, fatigue, fever, joint pains or swelling, photosensitivity, Raynaud's phenomenon, pleuritis, dry eyes, dry mouth) should be sought in any patient withILD. The CTDs may be difficult to rule out since the pulmonary manifestations occasionally precede the more typical systemic manifestations by months or years. The most common form of pulmonary involvement is a chronic interstitial pattern similar to that in patients withIPF. However, determining the precise nature of lung involvement in most of the CTDs is difficult due to the high incidence of lung involvement caused by disease-associated complications of esophageal dysfunction (predisposing to aspiration and secondary infections), respiratory muscle weakness (atelectasis and secondary infections), complications of therapy (opportunistic infections), and associated malignancies.

Progressive Systemic Sclerosis (PSS) (See alsoChap. 313) Clinical evidence ofILD is present in about one-half of patients with progressive systemic sclerosis, and pathologic evidence in three-quarters. Pulmonary function tests show a restrictive pattern and impaired diffusing capacity, often before any clinical or radiographic evidence of lung disease appears. Pulmonary vascular disease alone or in association with pulmonary fibrosis, pleuritis, or recurrent aspiration pneumonitis is strikingly resistant to current modes of therapy.

Rheumatoid Arthritis (RA) (See alsoChap. 312) ILDassociated with rheumatoid arthritis is more common in men. Pulmonary manifestations of rheumatoid arthritis include pleurisy with or without effusion, ILD in up to 20% of cases, necrobiotic nodules (nonpneumoconiotic intrapulmonary rheumatoid nodules) with or without cavities, Caplan's syndrome (rheumatoid pneumoconiosis), pulmonary hypertension secondary to rheumatoid pulmonary vasculitis,BOOP, and upper airway obstruction due to arytenoid arthritis.

Systemic Lupus Erythematosus (See also Chap. 311) Lung disease is a common complication inSLE. Pleuritis with or without effusion is the most common pulmonary manifestation. Other lung manifestations include the following: atelectasis, diaphragmatic dysfunction with loss of lung volumes, pulmonary vascular disease, pulmonary hemorrhage, uremic pulmonary edema, infectious pneumonia, andBOOP. Acute lupus pneumonitis characterized by pulmonary capillaritis leading to alveolar hemorrhage is common. Chronic, progressiveILD is uncommon. It is important to exclude pulmonary infection. Although pleuropulmonary involvement may not be evident clinically, pulmonary function testing, particularly D_{LCO}reveals abnormalities in many patients with SLE.

Polymyositis and Dermatomyositis (PM/DM) (See also Chap. 382) ILDoccurs in ~10% of patients with polymyositis and dermatomyositis, and the clinical features are similar to those ofIPF. Diffuse reticular or nodular opacities with or without an alveolar component occur radiographically, with a predilection for the lung bases. ILD occurs more commonly in the subgroup of patients with an anti-Jo-1 antibody that is directed to histidyl tRNA synthetase. Weakness of respiratory muscles contributing to aspiration pneumonia may be present. A rapidly progressive illness characterized by diffuse alveolar damage may cause respiratory failure.

Sjogren's Syndrome (See alsoChap. 314) General dryness and lack of airways secretion cause the major problems of hoarseness, cough, and bronchitis. Lymphocytic interstitial pneumonitis, lymphoma, pseudolymphoma, bronchiolitis, and bronchiolitis obliterans are associated with this condition. Lung biopsy is frequently required to establish a precise pulmonary diagnosis. Glucocorticoids have been used in the management ofILDassociated with Sjogren's syndrome with some degree of clinical success.

DRUG-INDUCED ILD (See alsoChap. 71)

Many classes of drugs have the potential to induce diffuseILD, which is manifest most commonly as exertional dyspnea and nonproductive cough. A detailed history of the medications taken by the patient is needed to identify drug-induced disease, including over-the-counter medications, oily nose drops, or petroleum products (mineral oil). In most cases, the pathogenesis is unknown, although a combination of direct toxic effects of the drug (or its metabolite) and indirect inflammatory and immunologic events is likely. The onset of the illness may be abrupt and fulminant, or it may be insidious, extending over weeks to months. The drug may have been taken for several years before a reaction develops (e.g., amiodarone), or the lung disease may occur weeks to years after the drug has been discontinued (e.g., carmustine). The extent and severity of disease are usually dose related. Treatment consists of discontinuation of any possible offending drug and supportive care.

CRYPTOGENIC ORGANIZING PNEUMONIA (COP)

Also known as idiopathicBOOP,COP is a clinicopathologic syndrome of unknown etiology. The onset is usually in the fifth and sixth decades. The presentation may be of a flu-like illness with cough, fever, malaise, fatigue, and weight loss. Inspiratory crackles are frequently present on examination. Pulmonary function is usually impaired, with a

restrictive defect and arterial hypoxemia being most common. The roentgenographic manifestations are distinctive, revealing bilateral, patchy, or diffuse alveolar opacities in the presence of normal lung volume. Recurrent and migratory pulmonary opacities are common.HRCTshows areas of air-space consolidation, ground-glass opacities, small nodular opacities and bronchial wall thickening and dilation. These changes occur more frequently in the periphery of the lung and in the lower lung zone. Lung biopsy shows granulation tissue within small airways, alveolar ducts, and airspaces, with chronic inflammation in the surrounding alveoli. Glucocorticoid therapy induces clinical recovery in two-thirds of patients. A few patients have rapidly progressive courses with fatal outcomes despite glucocorticoids.

Foci of organizing pneumonia (i.e., a "BOOPpattern") is a nonspecific reaction to lung injury found adjacent to other pathologic processes or as a component of other primary pulmonary disorders (e.g., cryptococcosis, Wegener's granulomatosis, lymphoma, hypersensitivity pneumonitis, and eosinophilic pneumonia). Consequently, the clinician must carefully reevaluate any patient found to have this histopathologic lesion to rule out these possibilities.

EOSINOPHILIC PNEUMONIA See Chap. 253

PULMONARY ALVEOLAR PROTEINOSIS

Although not strictly anILD, pulmonary alveolar proteinosis (PAP) resembles and is therefore considered with these conditions. It has been proposed that a defect in macrophage function, more specifically an impaired ability to process surfactant, may play a role in the pathogenesis of PAP. This diffuse disease is characterized by the accumulation of an amorphous, periodic acid-Schiff-positive lipoproteinaceous material in the distal air spaces. There is little or no lung inflammation, and the underlying lung architecture is preserved. Mutant mice lacking the gene for granulocyte-macrophage colony stimulating factor (GM-CSF) have a similar accumulation of surfactant and surfactant apoprotein in the alveolar spaces. Moreover, reconstitution of the respiratory epithelium of GM-CSF knockout mice with the GM-CSF gene completely corrects the alveolar proteinosis. Data from BAL studies in patients suggest that PAP is an autoimmune disease with neutralizing antibody of immunoglobulin G isotype against GM-CSF. These findings suggest that neutralization of GM-CSF bioactivity by the antibody causes dysfunction of alveolar macrophages, which results in reduced surfactant clearance.

The typical age of presentation is 30 to 50 years, and males predominate. The clinical presentation is usually insidious and manifested by progressive exertional dyspnea, fatigue, weight loss, and low-grade fever. A nonproductive cough is common, but occasionally expectoration of "chunky" gelatinous material may occur. Polycythemia, hypergammaglobulinemia, and increased LDH levels are frequent. Markedly elevated serum levels of lung surfactant proteins A and D have been found in PAP. Radiographically, bilateral symmetrical alveolar opacities located centrally in mid and lower lung zones result in a "bat-wing" distribution.HRCTshows a ground-glass opacification and thickened intralobular structures and interlobular septa. Whole lung lavage(s) through a double-lumen endotracheal tube provides relief to many patients with dyspnea or progressive hypoxemia and also may provide long-term benefit.

PULMONARY LYMPHANGIOLEIOMYOMATOSIS

PulmonaryLAM is a rare condition that afflicts premenopausal women and should be suspected in young women with emphysema, recurrent pneumothorax, or chylous pleural effusion. It is often misdiagnosed as asthma or chronic obstructive pulmonary disease. Pathologically, LAM is characterized by the proliferation of atypical pulmonary interstitial smooth muscle and cyst formation. The immature-appearing smooth-muscle cells react with monoclonal antibody HMB45, which recognizes a 100-kDa glycoprotein (gp100) originally found in human melanoma cells. Caucasians are affected much more commonly than members of other racial groups. The disease accelerates during pregnancy and abates after oopherectomy. Common complaints at presentation are dyspnea, cough, and chest pain. Hemoptysis may be life threatening. Spontaneous pneumothorax occurs in 50% of patients; it may be bilateral and necessitate pleurodesis. Chylothorax, chyloperitonium (chylous ascites), chyluria, and chylopericardium are other complications. Pulmonary function testing usually reveals an obstructive or mixed obstructive-restrictive pattern, and gas exchange is often abnormal.HRCTshows thin-walled cysts surrounded by normal lung without zonal predominance. Progression is common, with a median survival of 8 to 10 years from diagnosis. Oophorectomy, progesterone (10 mg/d), and, more recently, tamoxifen and luteinizing hormone-releasing hormone analogs have been used. Lung transplantation offers the only hope for cure despite reports of recurrent disease in the transplanted lung.

SYNDROMES OF ILD WITH DIFFUSE ALVEOLAR HEMORRHAGE

Injury to arterioles, venules, and the alveolar septal (alveolar wall or interstitial) capillaries can result in hemoptysis secondary to disruption of the alveolar-capillary basement membrane. This results in bleeding into the alveolar spaces, which characterizesDAH. Pulmonary capillaritis, characterized by a neutrophilic infiltration of the alveolar septae, may lead to necrosis of these structures, loss of capillary structural integrity, and the pouring of red blood cells into the alveolar space. Fibrinoid necrosis of the interstitium and red blood cells within the interstitial space are sometimes seen. Bland pulmonary hemorrhage (i.e., DAH without inflammation of the alveolar structures) may also occur.

The clinical onset is often abrupt, with cough, fever, and dyspnea. Severe respiratory distress requiring ventilatory support may be evident at initial presentation. Although hemoptysis is expected, it can be absent at the time of presentation in one-third of the cases. For patients without hemoptysis, new alveolar opacities, a falling hemoglobin level, and hemorrhagicBALfluid point to the diagnosis. The chest radiograph is nonspecific and most commonly shows new patchy or diffuse alveolar opacities. Recurrent episodes ofDAH may lead to pulmonary fibrosis, resulting in interstitial opacities on the chest radiograph. An elevated white blood cell count and falling hematocrit are frequent. Evidence for impaired renal function caused by focal segmental necrotizing glomerulonephritis, usually with crescent formation, may also be present.

Varying degrees of hypoxemia may occur and often are severe enough to require ventilatory support. TheD_{LCO}may be increased, resulting from the increased hemoglobin

within the alveoli compartment. Evaluation of either lung or renal tissue by immunofluorescent techniques indicates an absence of immune complexes (pauci-immune) in Wegener's granulomatosis, microscopic polyangiitis pauci-immune glomerulonephritis, and isolated pulmonary capillaritis. A granular pattern is found in theCTDs, particularlySLE, and a characteristic linear deposition is found in Goodpasture's syndrome. Granular deposition of IgA-containing immune complexes are present in Henoch-Schonlein purpura.

The mainstay of therapy for theDAHassociated with systemic vasculitis,CTD, Goodpasture's syndrome, and isolated pulmonary capillaritis is intravenous methylprenisolone, 0.5 to 2.0 g daily in divided doses for up to 5 days, followed by a gradual tapering, and then maintenance on an oral preparation. Prompt initiation of therapy is important, particularly in the face of renal insufficiency, since early initiation of therapy has the best chance of preserving renal function. The decision to start other immunosuppressive therapy (cyclophosphamide or azathioprine) acutely depends on the severity of illness.

Goodpasture's Syndrome Pulmonary hemorrhage and glomerulonephritis are features in most patients with this disease (Chap. 275). Autoantibodies to renal glomerular and lung alveolar basement membranes are present. This syndrome can present and recur asDAHwithout an associated glomerulonephritis. In such case, circulating anti-basement membrane antibody is often absent, and the only way to establish the diagnosis is by demonstrating linear immunofluorescence in lung tissue. The underlying histology may be bland hemorrhage or DAH associated with capillaritis. Plasmapheresis has been recommended as adjunctive treatment.

Idiopathic Pulmonary Hemosiderosis This condition is a diagnosis of exclusion. Only 20% of reported cases occur in adults. In children, the condition is associated with celiac disease, and elevated IgA levels are found in 50% of patients. These associations are lacking in most adults. A lung biopsy is usually necessary to document the lack of inflammatory injury in the lung tissues and to exclude other diseases with confidence.

INHERITED DISORDERS ASSOCIATED WITH ILD

Pulmonary opacities and respiratory symptoms typical ofILD can develop in related family members and in several inherited diseases. These include the phakomatoses, tuberous sclerosis and neurofibromatosis (Chap. 370), and the lysosomal storage diseases, Niemann-Pick disease and Gaucher's disease (Chap. 349). The Hermansky-Pudlak syndrome (Chap. 116) is an autosomal recessive disorder in which granulomatous colitis and ILD may occur. It is characterized by oculocutaneous albinism, bleeding diathesis secondary to platelet dysfunction, and the accumulation of a chromolipid, lipofuscin material in cells of the reticuloendothelial system. The pulmonary fibrosis is similar toIPF, but the alveolar macrophages may contain cytoplasmic ceroid-like inclusions.

ILD WITH A GRANULOMATOUS RESPONSE IN LUNG TISSUE OR VASCULAR STRUCTURES

Inhalation of organic dusts, which cause hypersensitivity pneumonitis, or of inorganic

dust, such as silica, which elicits a granulomatous inflammatory reaction leading toILD, produces diseases of known etiology (Table 259-1) that are discussed inChaps. 253 and254. Sarcoidosis (Chap. 318) is prominent among granulomatous diseases of unknown cause in which ILD is an important feature.

Pulmonary Langerhans Cell Histiocytosis (PLCH, Pulmonary Histiocytosis X, Langerhans Cell Granulomatosis, or Eosinophilic Granuloma PLCH is a rare, smoking-related, diffuse lung disease that primarily affects men between the ages of 20 and 40 years. The clinical presentation varies from an asymptomatic state to a rapidly progressive condition. The most common clinical manifestations at presentation are cough, dyspnea, chest pain, weight loss, and fever. Pneumothorax occurs in about 25% of patients. Hemoptysis and diabetes insipidus are rare manifestations. The radiographic features vary with the stage of the disease. The combination of ill-defined or stellate nodules (2 to 10 mm in diameter), reticular or nodular opacities, bizarre-shaped upper zone cysts, preservation of lung volume, and sparing of the costophrenic angles are characteristics of PLCH.HRCTthat reveals a combination of nodules and thin-walled cysts is virtually diagnostic of PLCH. The most frequent pulmonary function abnormality is a markedly reducedD_{LCO}, although varying degrees of restrictive disease, airflow limitation, and diminished exercise capacity may occur. Discontinuance of smoking is the key treatment, resulting in clinical improvement in one-third of patients. Most patients with PLCH suffer persistent or progressive disease. Death due to respiratory failure occurs in ~10% of patients.

Granulomatous Vasculitides (See also Chap. 317) The granulomatous vasculitides are characterized by pulmonary angiitis (i.e., inflammation and necrosis of blood vessels) with associated granuloma formation (i.e., infiltrates of lymphocytes, plasma cells, epithelioid cells, or histiocytes, with or without the presence of multinucleated giant cells, sometimes with tissue necrosis). The lungs are almost always involved, although any organ system may be affected. Wegener's granulomatosis and allergic angiitis and granulomatosis (Churg-Strauss syndrome) primarily affect the lung but are associated with a systemic vasculitis as well. The granulomatous vasculitides generally limited to the lung include necrotizing sarcoid granulomatosis and benign lymphocytic angiitis and granulomatosis. Granulomatous infection and pulmonary angiitis due to irritating embolic material (e.g., talc) are important known causes of pulmonary vasculitis.

LYMPHOCYTIC INFILTRATIVE DISORDERS

This group of disorders features lymphocyte and plasma cell infiltration of the lung parenchyma. The disorders either are benign or can behave as low-grade lymphomas. Included are angioimmunoblastic lymphadenopathy with dysproteinemia, a rare lymphoproliferative disorder characterized by diffuse lymphadenopathy, fever, hepatosplenomegaly, and hemolytic anemia, withILD in some cases.

Lymphocytic Interstitial Pneumonitis This rare form ofILDoccurs in adults, some of whom have an autoimmune disease or dysproteinemia. It has been reported in patients with Sjogren's syndrome and HIV infection.

Lymphomatoid Granulomatosis This multisystem disorder of unknown etiology is an angiocentric malignant (T cell) lymphoma characterized by a polymorphic lymphoid

infiltrate, an angiitis, and granulomatosis. Although it may affect virtually any organ, it is most frequently characterized by pulmonary, skin, and central nervous system involvement.

BRONCHOCENTRIC GRANULOMATOSIS

Rather than a specific clinical entity, bronchocentric granulomatosis (BG) is a descriptive histologic term that describes an uncommon and nonspecific pathologic response to a variety of airway injuries. There is evidence that BG is caused by a hypersensitivity reaction to *Aspergillus* or other fungi in patients with asthma. About half of the patients described have chronic asthma with severe wheezing and peripheral blood eosinophilia. In patients with asthma, BG probably represents one pathologic manifestation of allergic bronchopulmonary aspergillosis, or another allergic mycosis. In patients without asthma, BG has been associated with rheumatoid arthritis and a variety of infections, including tuberculosis, echinococcosis, histoplasmosis, coccidiodomycosis, and nocardiosis. The chest roentgenogram reveals irregularly shaped nodular or mass lesions with ill-defined margins, which are usually unilateral and solitary, with an upper-lobe predominance. Glucocorticoids are the treatment of choice, often with excellent outcome, although recurrences may occur as therapy is tapered or stopped.

(Bibliography omitted in Palm version)

Back to Table of Contents

260. PRIMARY PULMONARY HYPERTENSION - *Stuart Rich*

Primary pulmonary hypertension is an uncommon disease characterized by increased pulmonary artery pressure and pulmonary vascular resistance. The incidence has been estimated at approximately 2 cases per million. There is a female-to-male preponderance (1.7:1), with patients most commonly presenting in the third and fourth decades, although the age range is from infancy to greater than 60 years. Because the predominant symptom of primary pulmonary hypertension is dyspnea, which can have an insidious onset in an otherwise healthy person, the disease is typically diagnosed late in its course. By that time, the clinical and laboratory findings of severe pulmonary hypertension are usually present.

PATHOLOGY

The histopathology of primary pulmonary hypertension is not pathognomonic for the disease but represents a pulmonary arteriopathy that is observed in pulmonary hypertension from a variety of causes. A wide spectrum of vascular abnormalities involving the endothelium, smooth muscle cells, and extracellular matrix is present. Heterogeneity with respect to these abnormalities is often seen from patient to patient, and within patients. The most common features noted are medial hypertrophy, concentric and eccentric intimal fibrosis, recanalized thrombi appearing as fibrous webs, and plexiform lesions. In most patients, varying degrees of these abnormalities can be found. Rare variant forms of primary pulmonary hypertension also exist.

Pulmonary venoocclusive disease is a rare and distinct pathologic entity, found in fewer than 10% of patients with primary pulmonary hypertension. Histologically, it is manifest by widespread intimal proliferation and fibrosis of the intrapulmonary veins and venules, occasionally extending to the arteriolar bed. The pulmonary venous obstruction explains the increased pulmonary capillary wedge pressure observed in patients with advanced disease. These patients may develop orthopnea that can mimic left ventricular failure.

Pulmonary capillary hemangiomatosis is also a very rare form of primary pulmonary hypertension. Histologically, it is characterized by infiltrating thin-walled blood vessels that are widespread throughout the pulmonary interstitium and walls of the pulmonary arteries and veins. These patients often have hemoptysis as a clinical feature.

ETIOLOGY

It is likely that there are several pathobiologic processes that result in pulmonary hypertension as a final common pathway. These include inhibition of the voltage-regulated (Kv) potassium channel producing vasoconstriction secondary to contraction of the pulmonary artery smooth muscle cells, an imbalance in vasocontricting and vasodilating mediators that are involved in the control of pulmonary vascular tone (including prostacyclin and thromboxane), reduced expression of nitric oxide synthase in the endothelium of the pulmonary arterial bed, inflammation, thrombosis in situ of the pulmonary vascular bed from a procoagulant state, and persistent matrix protein synthesis in the pulmonary arteries. The types of abnormalities that occur are likely influenced by the patient's genotype and exposure to risk factors that serve to trigger these processes. Risk factors that have been linked to the

development of pulmonary hypertension include anorexigens, collagen vascular diseases, congenital systemic to pulmonary shunts, portal hypertension, and HIV infection.

Recently, a marked increase in the incidence of primary pulmonary hypertension occurred in Europe and the United States as a result of the widespread use of the fenfluramine appetite suppressants. The clinical and pathologic features of these cases were identical to patients with primary pulmonary hypertension who were unexposed. Atlhough very limited exposure to the fenfluramines can cause primary pulmonary hypertension, the risk increased dramatically with prolonged use. Like the experience with aminorex, an anorexigen that produced a similar epidemic in the 1960s, the incidence of primary pulmonary hypertension fell when the drugs were withdrawn from the market. The mechanism by which these agents produce pulmonary hypertension is unknown.

GENETIC CONSIDERATIONS

The locus of a gene linked to familial primary pulmonary hypertension has been identified on chromosome 2q31-32. Familial primary pulmonary hypertension occurs in approximately 6 to 12% of cases and is characterized by autosomal dominant inheritance, variable age of onset, and incomplete penetrance. The clinical and pathologic features of familial and sporadic primary pulmonary hypertension are virtually identical. Genetic anticipation, which relates to offspring of subsequent generations manifesting the disease at younger ages or with greater severity, is also a feature. Trinucleotide repeat expansion, originally described in several neurologic disorders, remains the only biologic explanation for genetic anticipation and raises the possibility that the pathogenesis of familial primary pulmonary hypertension may have a neurologic basis. Patients who present with sporadic disease probably possess a genetic predisposition that becomes expressed following exposure to an external trigger or risk factor.

PATHOPHYSIOLOGY

The underlying hemodynamic derangement in primary pulmonary hypertension is an increased resistance to pulmonary blood flow. Early in the disease there is a marked elevation in pulmonary artery pressure with relatively normal cardiac function. Over time the cardiac output becomes progressively reduced rather than the pulmonary artery pressure becoming progressively increased. Initially, the pulmonary arteries may respond to vasodilators, but as the disease progresses, the elevated pulmonary vascular resistance becomes fixed. The pulmonary capillary wedge pressure remains normal until the late stages, when it tends to rise in response to impaired diastolic filling of the left ventricle due to the altered configuration of the intraventricular septum. Eventually, as the right ventricle fails, the right atrial and right ventricular end-diastolic pressures rise in an attempt to compensate for the myocardial depression that has developed in response to chronic severe right ventricular pressure overload.

Pulmonary function is usually normal in primary pulmonary hypertension, although a mild restrictive pattern (Chap. 250) is sometimes seen. Hypoxemia is common and is believed to be due to a mismatch between pulmonary ventilation and perfusion,

magnified by a low cardiac output. Occasional patients with a patent foramen ovale may develop right-to-left shunting, which can also contribute to systemic arterial desaturation.

DIAGNOSIS

Primary pulmonary hypertension refers to pulmonary arterial hypertension wihtout an identifiable risk factor. Clinically, primary pulmonary arterial hypertension should be distinguished from pulmonary venous hypertension, pulmonary hypertension associated with disorders of the respiratory system and/or hypoxema, and pulmonary hypertension due to chronic thrombotic and/or embolic disease (Chap. 261).

A thorough diagnostic evaluation to look for all potential causes should be undertaken (Fig. 260-1). The history usually reveals the gradual onset of shortness of breath with effort, progressing until the patient is dyspneic with minimal activity. The average duration from symptom onset until diagnosis is 2.5 years. Other common symptoms are fatigue, angina pectoris that likely represents right ventricular ischemia, syncope, near syncope, and peripheral edema.

The physical examination is characteristic. Increased jugular venous pressure, a reduced carotid pulse, and an easily palpable right ventricular lift are typical. Most patients have an increased pulmonic component of the second heart sound and right-sided third and fourth heart sounds. Tricuspid regurgitation is a clinical feature of right ventricular failure. Peripheral cyanosis and/or edema tend to occur in later stages of the disease. Clubbing is not a feature.

The chest x-ray generally shows enlarged central pulmonary arteries and clear lung fields. The electrocardiogram usually reveals right axis deviation and right ventricular hypertrophy. The echocardiogram demonstrates right ventricular enlargement, a reduction in left ventricular cavity size, and abnormal septal configuration consistent with right ventricular pressure overload. Doppler studies have revealed a marked dependence on atrial systole for ventricular filling. Hypoxemia, hypocapnia, and an abnormal diffusing capacity for carbon monoxide are almost invariable findings. A mild restrictive pattern on pulmonary function is sometimes observed, but evidence of airways obstruction suggests a secondary etiology for the pulmonary hypertension. The presence of significant restrictive changes on pulmonary function testing (Chap. 250) should prompt a high-resolution computed tomographic scan to look for interstitial lung disease, which may otherwise not be obvious. A perfusion lung scan may be normal or abnormal with multiple diffuse patchy filling defects of a nonsegmental nature and not suggestive of pulmonary thromboembolism. If the lung scan reveals perfusion defects of a segmental or subsegmental nature, a pulmonary angiogram must be done. Severe pulmonary hypertension in a patient with a high-probability lung scan should suggest a chronic process and *not* acute pulmonary embolism, since the nonconditioned right ventricle is unable to generate high systolic pressures acutely in the face of pulmonary thromboembolism. Chronic thromboembolic obstruction of the large pulmonary arteries (Chap. 261) can mimic primary pulmonary hypertension but can be amenable to treatment with surgical thromboendarterectomy.

There is risk in performing pulmonary angiography in patients with primary pulmonary

hypertension, and it is recommended that selective or subselective injections with small amounts of low-osmolar, nonionic contrast material be made following the pretreatment with 1 mg atropine to prevent vagally mediated bradycardia.

Cardiac catheterization is mandatory to characterize the disease and exclude an underlying cardiac shunt as the cause. The use of balloon-flotation catheters, especially those with removable guidewires, can facilitate right heart catheterization. A right-to-left shunt might be attributable to a patent foramen ovale, but any left-to-right shunting implies the presence of a congenital defect. Although it may be difficult to obtain, the pulmonary capillary wedge pressure is normal. If it is increased, left heart catheterization should also be performed to exclude mitral stenosis or increased left ventricular end-diastolic pressures as the cause of the pulmonary hypertension. Although the diagnostic evaluation of these patients can be hazardous, experience from a national multicenter study revealed no mortality or serious morbidity in more than 300 patients whose evaluation included pulmonary angiography and cardiac catheterization. It is not necessary to perform an open lung biopsy in these patients to make an accurate diagnosis. Laboratory tests should also be performed, including antinuclear antibody and HIV testing.

On occasion, a patient may have marked elevations in pulmonary artery pressure in association with mild obstructive or interstitial lung disease, essential hypertension, ischemic heart disease, or valvular heart disease. Although it may appear that the pulmonary hypertension is out of proportion to the underlying associated condition, it likely represents a pulmonary vasoconstrictive response to the associated condition, which is serving as a trigger of pulmonary arterial hypertension. Thus severe pulmonary hypertension can coexist with mild chronic obstructive pulmonary disease, small intracardiac shunts, mild mitral stenosis, and even ischemic heart disease. The distinction is important because the treatment of pulmonary hypertension should always include treating the underlying associated cause.

NATURAL HISTORY

The natural history of primary pulmonary hypertension is unknown because initially the disease is largely asymptomatic. Several older series have reported a mean survival of 2 to 3 years for patients from the time of diagnosis. Functional class is a strong predictor of survival, since patients who are New York Heart Association functional classes II and III have a mean survival of 3.5 years compared with those who are functional class IV, in whom the mean survival is 6 months. The cause of death is usually right ventricular failure or sudden death; sudden death appears to be a late feature of the disease. Increased right atrial pressure above 15 mmHg and reduced cardiac index below 2 (L/min)/$m^2$are hemodynamic predictors of a poor prognosis.

TREATMENT

Because the pulmonary vascular resistance increases dramatically with exercise, patients should be cautioned against participating in activities that demand increased physical stress. Digoxin may increase cardiac output and lower circulating levels of norepinephrine. Diuretic therapy relieves dyspnea and peripheral edema and may be useful in reducing right ventricular volume overload in the presence of tricuspid

regurgitation.

It is recommended that all patients in whom primary pulmonary hypertension is confirmed undergo acute drug testing with short-acting pulmonary vasodilators to determine the extent of pulmonary vasodilator reserve or reactivity (Fig. 260-2). Intravenous adenosine, inhaled nitric oxide, and intravenous prostacyclin all appear to have similar effects in reducing pulmonary vascular resistance acutely with little effect on the systemic vascular bed. Adenosine is given as a constant infusion in doses of 50 (ug/kg)/min and increased every 2 min until side effects develop. Similarly, prostacyclin is given in doses of 2 (ng/kg)/min and increased every 30 min until side effects develop. Maximal physiologic effectiveness of the therapy is determined at the highest tolerated dose. Nitric oxide is generally administered via inhalation in 5 to 10 parts per million and increased every few minutes until no further effectiveness is obtained.

Calcium Channel Antagonists Patients who have substantial reductions in pulmonary vascular resistance from the short-acting vasodilators may be candidates to receive oral calcium channel blockers. These drugs should be administered under direct hemodynamic guidance in order to determine effectiveness and safety. Typically, patients will require high doses (e.g., nifedipine, 120 to 240 mg/d, or diltiazem, 540 to 900 mg/d).*

*These agents have not been approved for the treatment of primary pulmonary hypertension by the U.S. Food and Drug Administration.

Patients who manifest significant reductions in mean pulmonary artery pressure and pulmonary vascular resistance should demonstrate improved symptoms, regression of right ventricular hypertrophy, and improved survival with chronic therapy. However, fewer than half the patients who are responsive to the short-acting vasodilators will respond to this regimen. It is unknown whether the response to calcium blockers depends on the histologic subtype, but the therapy appears to be more successful in patients who are diagnosed early and have less advanced disease.

Prostacyclin This agent has been approved as a treatment of primary pulmonary hypertension for patients who are functional class III or IV and unresponsive to conventional therapy. Clinical trials have demonstrated that patients realize an improvement in symptoms and exercise tolerance and reduction in mortality, even if no acute hemodynamic response to drug challenge occurs. The drug can only be administered intravenously and requires placement of a permanent central venous catheter and continuous dose titration, as tolerance develops in all patients over a short period of time. The optimal dose has not been determined. Patients may deteriorate clinically from too much or too little drug. The side effects of prostacyclin, which include flushing, jaw pain, and diarrhea, are generally tolerated by most patients. The major problems with this therapy have been infections related to the venous catheter, which requires close monitoring and diligence on behalf of the patient. Recent data suggest that prostacyclin, in addition to its vasodilator and antithrombotic properties, may lead to reversal of the vascular remodeling that occurs in primary pulmonary hypertension. Long-term use of prostacyclin has been associated with adverse effects such as severe thrombocytopenia and severe foot pain, which can be disabling. The basis for these conditions is unknown. Because of the complexity involved in managing patients on

prostacyclin, it has been recommended that they be referred to centers with expertise in managing primary pulmonary hypertension for initiation of therapy.

Adverse Effects The administration of vasodilators can have serious acute and chronic adverse effects. The most common response is a reduction in pulmonary vascular resistance, manifest by an increased cardiac output, without a reduction in the mean pulmonary artery pressure. This results in increased stroke work of the right ventricle, which can result in worsening of ventricular function and precipitate right ventricular failure over time. In addition, maintenance of adequate systemic blood pressure is crucial, since right ventricular coronary blood flow is already compromised due to the loss of the normal gradient for myocardial perfusion between the aorta and right ventricle. Vasodilator drugs can provoke acute right ventricular ischemia, and deaths have been reported. For these reasons, the pharmacologic evaluation of primary pulmonary hypertension should always be undertaken with direct monitoring of systemic and pulmonary arterial pressures and cardiac output.

Anticoagulant therapy has also been advocated based on the evidence that thrombosis in situ is common. One retrospective study and one prospective study have demonstrated that the anticoagulant warfarin increases the survival of patients with primary pulmonary hypertension, and thus consideration for its use should be given to all patients. The dose of warfarin is generally titrated to achieve an increase in INR of 2.0 to 2.5 of control. Anticoagulants should not be expected to cause regression of the disease and result in any substantial change in symptoms.

Transplantation Because of the dramatic effects that prostacyclin has had in stabilizing the clinical course of patients with advanced disease, transplantation should be considered for patients on prostacyclin who develop or continue to manifest right heart failure. Acceptable results have been achieved with heart-lung, bilateral lung, and single lung transplant (Chap. 267). The availability of donor organs often influences the choice of procedure. The operation is best reserved for patients who are in the advanced stages of the disease in spite of medical therapy, or in whom medical therapy is not tolerated. Recurrence of disease has not been reported in any patient with primary pulmonary hypertension who has undergone single lung or heart-lung transplantation.

(Bibliography omitted in Palm version)

Back to Table of Contents

261. PULMONARY THROMBOEMBOLISM - *Samuel Z. Goldhaber*

GENETIC CONSIDERATIONS

Rudolf Virchow postulated more than a century ago that three potentially overlapping factors predisposed to venous thrombosis: (1) local trauma to the vessel wall; (2) hypercoagulability; and (3) stasis. We now believe that many patients who suffer pulmonary thromboembolism (PTE) have an underlying inherited predisposition that remains clinically silent until an acquired stressor occurs such as surgery, obesity, or pregnancy (Table 261-1). When PTE is identified, a detailed family history for venous thromboembolism should be obtained.

Factor V Leiden The most frequent inherited predisposition to hypercoagulability is resistance to the endogenous anticoagulant protein, activated protein C. The phenotype of activated protein C resistance is associated with a single point mutation, designated *factor V Leiden*, in the factor V gene. This missense mutation -- a single nucleotide substitution of adenine for guanine 1691 -- causes an amino acid substitution of glutamine for arginine at position 506.

The prevalence of the heterozygous state was about 6% in healthy American male physicians participating in the Physicians' Health Study and was three times higher among those physicians who subsequently developed venous thrombosis. Furthermore, after anticoagulation (for at least 3 months) was completed and discontinued, those participants with factor V Leiden had a much higher rate of recurrent venous thrombosis than those without. A single-point mutation in the 3¢untranslated region of the prothrombin gene (G-to-A transition at nucleotide position 20210) appears to be associated with increased levels of prothrombin (factor II), the precursor of thrombin. In the Physicians' Health Study, the prevalence of the prothrombin gene mutation among control subjects was 3.9%. The G20210A mutation conferred an approximate doubling of the risk of venous thrombosis. Nevertheless, factor V Leiden is more common than all other (identified) inherited hypercoagulable states, including the prothrombin gene mutation, deficiencies in protein C, protein S, antithrombin III, and disorders of plasminogen (Chap. 117).

PATHOPHYSIOLOGY

EMBOLIZATION

When venous thrombi become dislodged from their site of formation, they embolize to the pulmonary arterial circulation or, paradoxically, to the arterial circulation through a patent foramen ovale or atrial septal defect. About half of patients with pelvic vein thrombosis or proximal leg deep venous thrombosis (DVT) havePTE, which is usually asymptomatic. Isolated calf vein or upper extremity venous thromboses also pose a risk (albeit lower) of PTE. Isolated calf vein thrombi are the most common source of paradoxical embolism.

PHYSIOLOGY

Pulmonary embolism can have the following effects:

1. *Increased pulmonary vascular resistance* due to vascular obstruction or neurohumoral agents including serotonin

2. *Impaired gas exchange* due to increased alveolar dead space from vascular obstruction and hypoxemia from alveolar hypoventilation in the nonobstructed lung, right-to-left shunting, and impaired carbon monoxide transfer due to loss of gas exchange surface

3. *Alveolar hyperventilation* due to reflex stimulation of irritant receptors

4. *Increased airway resistance* due to bronchoconstriction

5. *Decreased pulmonary compliance* due to lung edema, lung hemorrhage, or loss of surfactant

Right Ventricular Dysfunction Progressive right heart failure is the usual cause of death fromPTE. In the International Cooperative Pulmonary Embolism Registry (ICOPER), the presence of right ventricular dysfunction on baseline echocardiography of PTE patients was associated with a doubling of the 3-month mortality rate. As pulmonary vascular resistance increases, right ventricular wall tension rises and perpetuates further right ventricular dilatation and dysfunction. Consequently, the interventricular septum bulges into and compresses an intrinsically normal left ventricle. Increased right ventricular wall tension also compresses the right coronary artery and may precipitate myocardial ischemia and right ventricular infarction. Underfilling of the left ventricle may lead to a fall in left ventricular output and systemic arterial pressure, thereby provoking myocardial ischemia due to compromised coronary artery perfusion. Eventually, circulatory collapse and death may ensue.

DIAGNOSIS

The clinical setting can be immensely helpful in suggesting the diagnosis ofPTE. Patients with prior venous thromboembolism are at increased risk of recurrence (Table 261-1).

CLINICAL SYNDROMES

Patients with *massivePTE*present with systemic arterial hypotension and usually have anatomically widespread thromboembolism. Primary therapy with thrombolysis or embolectomy offers the greatest chance of survival. Those with *moderate to large PTE* have right ventricular hypokinesis on echocardiography but normal systemic arterial pressure. Optimal management is controversial; such patients may benefit from primary therapy to prevent recurrent embolism. Patients with *small to moderate PTE* have both normal right heart function and normal systemic arterial pressure. They have a good prognosis with either adequate anticoagulation or an inferior vena caval filter. The presence of *pulmonary infarction* usually indicates a small PTE, but one that is exquisitely painful, because it lodges near the innervation of pleural nerves.

Nonthrombotic pulmonary embolism may be easily overlooked. Possible etiologies

include fat embolism after blunt trauma and long bone fractures, tumor embolism, or air embolism. Intravenous drug users may inject themselves with a wide array of substances, such as hair, talc, or cotton. *Amniotic fluid embolism* occurs when fetal membranes leak or tear at the placental margin. The pulmonary edema seen in this syndrome is probably due primarily to alveolar capillary leakage.

SYMPTOMS AND SIGNS

Dyspnea is the most frequent symptom of PTE, and tachypnea is its most frequent sign. Whereas dyspnea, syncope, hypotension, or cyanosis indicate a massive PTE, pleuritic pain, cough, or hemoptysis often suggest a small embolism located distally near the pleura. On *physical examination*, young and previously healthy individuals may simply appear anxious but otherwise seem deceptively well, even with an anatomically large PTE. They need not have "classic" signs such as tachycardia, low-grade fever, neck vein distention, or an accentuated pulmonic component of the second heart sound. Sometimes, a paradoxical bradycardia occurs.

In older patients who complain of vague chest discomfort, the diagnosis ofPTE may not be apparent unless signs of right heart failure are present. Unfortunately, because acute coronary ischemic syndromes are so common, one may overlook the possibility of life-threatening PTE and may inadvertently discharge these patients from the hospital after the exclusion of myocardial infarction with serial cardiac enzyme measurements and electrocardiograms.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis ofPTE is broad (Table 261-2). Although PTE is known as "the great masquerader," quite often another illness simulates PTE. For example, when the proposed diagnosis of PTE is supposedly confirmed with a combination of dyspnea, chest pain, and an abnormal lung scan, the correct diagnosis of pneumonia might become apparent 12 h later when an infiltrate blossoms on chest x-ray, purulent sputum is first produced, and high fever and shaking chills develop.

Some patients havePTE and a coexisting illness such as pneumonia or heart failure. In such circumstances, clinical improvement will often fail to occur despite standard medical treatment of the concomitant illness. This situation can serve as a clinical clue to the possible coexistence of PTE.

NONIMAGING DIAGNOSTIC MODALITIES

These are generally safer, less expensive, but also less specific than diagnostic modalities that employ imaging.

Blood Tests The quantitative *plasma D*-dimer *enzyme-linked immunosorbent assay (ELISA)* level is elevated (>500 ng/mL) in more than 90% of patients withPTE, reflecting plasmin's breakdown of fibrin and indicating endogenous (though clinically ineffective) thrombolysis. A qualitative latex agglutination D-dimer assay, which is more readily available and less expensive than an ELISA, can be obtained initially; if elevated, the ELISA will also be elevated. However, if the latex agglutination is normal, a D-dimer

ELISA should be obtained, because the ELISA is much more sensitive than the latex agglutination D-dimer assay, which cannot be used to exclude PTE. The plasma D-dimer ELISA has a high negative predictive value and can be used to help exclude PTE. However, neither D-dimer assay is specific. Levels increase in patients with myocardial infarction, sepsis, or almost any systemic illness.

Data from the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) indicate that, contrary to classic teaching, *arterial blood gases* lack diagnostic utility forPTE. Among patients suspected of PTE, neither the room air arterial P_{O2} nor calculation of the alveolar-arterial oxygen gradient can reliably differentiate or triage patients who actually have PTE at angiography.

Electrocardiogram Classic abnormalities include sinus tachycardia; new-onset atrial fibrillation or flutter; and an S wave in lead I, a Q wave in lead III, and an inverted T wave in lead III (Chap. 226). Often, the QRS axis is greater than 90°. T-wave inversion in leads V_1 to V_4reflects right ventricular strain.

NONINVASIVE IMAGING MODALITIES

Chest Roentgenography A normal or near-normal chest x-ray in a dyspneic patient suggestsPTE. Well-established abnormalities include focal oligemia (Westermark's sign), a peripheral wedged-shaped density above the diaphragm (Hampton's hump), or an enlarged right descending pulmonary artery (Palla's sign).

Venous Ultrasonography ConfirmedDVT is usually an adequate surrogate forPTE. Ultrasonography of the deep venous system relies upon loss of vein compressibility as the primary criterion for DVT. About one-third of patients with PTE have no imaging evidence of DVT. In these situations, the clot may have already embolized to the lung or is in the pelvic veins, where ultrasonography is usually inadequate. Therefore, the workup for PTE should continue if there is high clinical suspicion, despite a normal ultrasound examination.

Lung Scanning (See alsoChap. 251) Lung scanning is the principal imaging test for the diagnosis ofPTE. Small particulate aggregates of albumin labeled with a gamma-emitting radionuclide are injected intravenously and are trapped in the pulmonary capillary bed. A perfusion scan defect indicates absent or decreased blood flow, possibly due to PTE. Ventilation scans, obtained with radiolabeled inhaled gases such as xenon or krypton, improve the specificity of the perfusion scan. Abnormal ventilation scans indicate abnormal nonventilated lung, thereby providing possible explanations for perfusion defects other than acute PTE. A high probability scan for PTE is defined as having two or more segmental perfusion defects in the presence of normal ventilation (Fig. 261-1).

Lung scanning is particularly useful if the results are normal or near-normal, or if there is a high probability forPTE. The diagnosis of PTE is very unlikely in patients with normal and near-normal scans but, in contrast, is about 90% certain in patients with high-probability scans. Unfortunately, fewer than half of patients with angiographically confirmed PTE have a high-probability scan. Importantly, as many as 40% of patients with high clinical suspicion for PTE and "low-probability" scans do, in fact, have PTE at

angiography.

Chest CT Computed tomography (CT) of the chest with intravenous contrast effectively diagnoses large, centralPTE but may fail to detect more peripherally located thrombi that are clinically important. In a comparison with standard contrast pulmonary angiography at Massachusetts General Hospital, the sensitivity of chest CT for PTE was only 60%.

Echocardiography This technique is useful for rapid triage of acutely ill patients who may havePTE. Bedside echocardiography can usually reliably differentiate among illnesses that have radically different treatment, including acute myocardial infarction, pericardial tamponade, dissection of the aorta, and PTE complicated by right heart failure. Detection of right ventricular dysfunction due to PTE helps to stratify the risk, delineate the prognosis, and plan optimal management.

INVASIVE DIAGNOSTIC MODALITIES

Pulmonary Angiography Selective pulmonary angiography is the most specific examination available for establishing the definitive diagnosis ofPTE and can detect emboli as small as 1 to 2 mm. A definitive diagnosis of PTE depends upon visualization of an intraluminal filling defect in more than one projection. Secondary signs of PTE include abrupt occlusion ("cut-off") of vessels; segmental oligemia or avascularity; a prolonged arterial phase with slow filling; or tortuous, tapering peripheral vessels.

Pulmonary angiography can be carried out safely among properly selected patients at hospitals that perform at least several studies per month. InPIOPED, the procedure resulted in death in five patients (0.5%), two of whom had severe heart failure prior to the procedure. Angiography is most useful when the clinical likelihood ofPTEdiffers substantially from the lung scan result or when the lung scan is of intermediate probability for PTE.

Contrast Phlebography This technique has been mostly replaced by ultrasonography. Venography is costly, uncomfortable, and occasionally results in contrast allergy or contrast-induced phlebitis. Contrast phlebography is worthwhile when there is a discrepancy between the clinical suspicion and the ultrasound result. Phlebography is also useful for diagnosing isolated calf vein thrombosis or recurrentDVT. A recently approved nuclear medicine test utilizing a synthetic peptide that binds preferentially to the glycoprotein IIb/IIIa receptors on activated platelets may eventually replace contrast phlebography in clinical practice. This radiopharmaceutical permits scintigraphic imaging of acute DVT and may be especially useful for differentiating acute from chronic DVT.

INTEGRATED DIAGNOSTIC APPROACH

We advocate an integrated diagnostic approach to streamline the workup ofPTE(Fig. 261-2). This strategy combines the clinical likelihood of PTE with the results of noninvasive testing especially D-dimerELISA, venous ultrasonography, and lung scanning to determine whether pulmonary angiography is warranted.

TREATMENT

ConsensusGuidelinesfrom the American College of Chest Physicians are summarized as follows.

Primary versus Secondary Therapy Primary therapy consists of clot dissolution with thrombolysis or removal ofPTE by embolectomy. Anticoagulation with heparin and warfarin or placement of an inferior vena caval filter constitutes secondary prevention of recurrent PTE rather than primary therapy.

Primary therapy should be reserved for patients at high risk of an adverse clinical outcome. When right ventricular function remains normal, patients typically have good clinical outcomes with anticoagulation alone (Fig. 261-3).

Adjunctive Therapy Important adjunctive measures include pain relief (especially with nonsteroidal anti-inflammatory agents), supplemental oxygenation, and psychological support. Dobutamine -- ab-adrenergic agonist with positive inotropic and pulmonary vasolidating effects -- may successfully treat right heart failure and cardiogenic shock. Volume loading should be undertaken cautiously because increased right ventricular dilatation can lead to even further reductions in left ventricular forward output.

Heparin Heparin binds to and accelerates the activity of antithrombin III, an enzyme that inhibits the coagulation factors thrombin (factor IIa), Xa, IXa, XIa, and XIIa. Heparin thus prevents additional thrombus formation and permits endogenous fibrinolytic mechanisms to lyse clot that has already formed. After 5 to 7 days of heparin, residual thrombus begins to stabilize in the endothelium of the vein or pulmonary artery. However, heparin does *not* directly dissolve thrombus that already exists.

Low-Molecular-Weight Heparins These fragments of unfractionated heparin exhibit less binding to plasma proteins and endothelial cells and consequently have greater bioavailability, a more predictable dose response, and a longer half-life than unfractionated heparin. No laboratory monitoring or dose adjustment is needed unless the patient is markedly obese or has renal insufficiency. Therefore, low-molecular-weight heparins are far more convenient to use than unfractionated heparin.

A meta-analysis of more than 3,500 acuteDVTpatients showed that those treated with low-molecular-weight heparin had an overall 29% reduction in mortality and major bleeding compared with the unfractionated heparin group. *Enoxaparin*, originally approved for prophylaxis, has recently received Food and Drug Administration approval for treatment ofPTE in the presence of DVT with a once-daily dose of 1.5 mg/kg subcutaneously. However, it is almost always administered as 1 mg/kg twice daily. *Dalteparin* is approved for prophylaxis but not for treatment of venous thromboembolism.

Dosing For unfractionated heparin, a typical bolus is 5000 to 10,000 units followed by a continuous infusion of 1000 to 1500 units/h. An activated partial thromboplastin time that is at least twice the control value should provide a therapeutic level of heparin. Nomograms based upon a patient's weight may assist in adjusting the infusion rate of heparin.

Complications The most important adverse effect of heparin is hemorrhage. For life-threatening or intracranial hemorrhage, protamine sulfate can be administered. Heparin-associated thrombocytopenia and osteopenia are far less common with low-molecular-weight heparins than with unfractionated heparin. Heparin-associated elevations in transaminase levels occur commonly but are rarely associated with clinical toxicity.

Warfarin This vitamin K antagonist prevents g carboxylation activation of coagulation factors II, VII, IX, and X. The full effect of warfarin often requires 5 days, even if the prothrombin time, used for monitoring, becomes elevated more rapidly. When warfarin is initiated during an active thrombotic state, the levels of protein C and S decline, thus creating a thrombogenic potential. By overlapping heparin and warfarin for 5 days, the procoagulant effect of unopposed warfarin can be counteracted. Thus, heparin acts as a "bridge" until the full anticoagulant effect of warfarin is obtained.

Dosing In an average-sized adult, warfarin is usually initiated in a dose of 5 mg. Doses of 7.5 or 10 mg can be used in obese or large framed young patients who are otherwise healthy. Patients who are malnourished or who have received prolonged courses of antibiotics are probably deficient in vitamin K and should receive smaller initial doses of warfarin, such as 2.5 mg. The prothrombin time is standardized by using the International Normalized Ratio (INR) to assess the anticoagulant effect of warfarin (Chap. 118). The target INR should be approximately 2.5-3.0.

Complications As with heparin, bleeding is the most important and common complication associated with warfarin administration. Life-threatening bleeding can be treated with cryoprecipitate or fresh frozen plasma (usually 2 units) to achieve immediate hemostasis. For less serious bleeding, or an excessively highINR in the absence of bleeding, vitamin K may be administered. An initial dose of 5 to 10 mg subcutaneously will help lower the INR toward the upper portion of the therapeutic range within about 6 h. Reversing excessive INRs with oral rather than subcutaneous vitamin K will facilitate re-establishing a stable dose of warfarin.

Warfarin-induced skin necrosis is a rare complication that may be related to warfarin-induced reduction of protein C. It is usually associated with administration of a high initial dose of warfarin during an acute thrombotic state in which heparin is withheld. During pregnancy, warfarin should be avoided if possible because of warfarin embryopathy, which is most common with exposure during the sixth through twelfth weeks of gestation. However, women can take warfarin postpartum and breast feed safely.

Duration of Anticoagulation After discontinuation of anticoagulation, the risk of recurrentPTE is surprisingly high. Nevertheless, the optimal duration of anticoagulation remains unknown. Schulman and colleagues found that after a 6-month course of anticoagulation, 14% of PTE patients suffered a recurrent venous thromboembolism within the ensuing 2 years. The recurrence rate was twice as high among patients who received only 6 weeks of anticoagulation. It is reasonable to anticoagulate the first episode of PTE for at least 6 months.

Inferior Vena Caval Filters When anticoagulation cannot be undertaken because of active bleeding, insertion of an inferior vena caval filter is usually necessary. Other indications include recurrent venous thrombosis despite adequate anticoagulation, prevention of recurrentPTE in patients with right heart failure who are not candidates for thrombolysis, or prophylaxis of extremely high risk patients. The Bird's Nest filter infrarenally or, if necessary, a Greenfield filter suprarenally are recommended.

Thrombolysis Thrombolytic therapy may rapidly reverse right heart failure and thus lead to a lower rate of death and recurrentPTE. Thrombolysis usually achieves the following: (1) dissolves much of the anatomically obstructing pulmonary arterial thrombus; (2) prevents the continued release of serotonin and other neurohumoral factors that might otherwise exacerbate pulmonary hypertension; and (3) dissolves much of the source of the thrombus in the pelvic or deep leg veins, thereby decreasing the likelihood of recurrent PTE.

The preferred thrombolytic regimen is 100 mg of recombinant tissue plasminogen activator administered as a continuous peripheral intravenous infusion over 2 h. Patients appear to respond to thrombolysis for up to 14 days after thePTEoccurred.

Contraindications to thrombolysis include intracranial disease, recent surgery, or trauma. There is about a 1% risk of intracranial hemorrhage. Careful screening of patients for contraindications to thrombolysis is the best way to minimize bleeding risk.

Pulmonary Thromboendarterectomy Patients who develop chronic pulmonary hypertension due to priorPTE may become severely dyspneic at rest or with minimal exertion. They should be considered for pulmonary thromboendarterectomy which, if successful, can markedly reduce and at times even cure pulmonary hypertension.

Prevention Prevention ofPTE is of paramount importance because it is both difficult to recognize and expensive to treat. Fortunately, effective mechanical and pharmacologic prophylaxis modalities are widely available and usually effective (Table 261-3).

(Bibliography omitted in Palm version)

Back to Table of Contents

262. DISORDERS OF THE PLEURA, MEDIASTINUM, AND DIAPHRAGM - *Richard W. Light*

DISORDERS OF THE PLEURA

PLEURAL EFFUSION

The pleural space lies between the lung and chest wall and normally contains a very thin layer of fluid, which serves as a coupling system. A pleural effusion is present when there is an excess quantity of fluid in the pleural space.

Etiology Pleural fluid accumulates when pleural fluid formation exceeds pleural fluid absorption. Normally, fluid enters the pleural space from the capillaries in the parietal pleura and is removed via the lymphatics situated in the parietal pleura. Fluid also can enter the pleural space from the interstitial spaces of the lung via the visceral pleura or from the peritoneal cavity via small holes in the diaphragm. The lymphatics have the capacity to absorb 20 times more fluid than is normally formed. Accordingly, a pleural effusion may develop when there is excess pleural fluid formation (from the parietal pleura, the interstitial spaces of the lung, or the peritoneal cavity) or when there is decreased fluid removal by the lymphatics.

Diagnostic Approach When a patient is found to have a pleural effusion, an effort should be made to determine the cause (Fig. 262-1). The first step is to determine whether the effusion is a transudate or an exudate. A *transudative* pleural effusion occurs when *systemic factors* that influence the formation and absorption of pleural fluid are altered. The leading causes of transudative pleural effusions in the United States are left ventricular failure, pulmonary embolism, and cirrhosis. An *exudative* pleural effusion occurs when *local factors* that influence the formation and absorption of pleural fluid are altered. The leading causes of exudative pleural effusions are bacterial pneumonia, malignancy, viral infection, and pulmonary embolism. The primary reason to make this differentiation is that additional diagnostic procedures are indicated with exudative effusions to define the cause of the local disease.

Transudative and exudative pleural effusions are distinguished by measuring the lactate dehydrogenase (LDH) and protein levels in the pleural fluid. Exudative pleural effusions meet at least one of the following criteria, whereas transudative pleural effusions meet none:

1. pleural fluid protein/serum protein>0.5

2. pleural fluid LDH/serum LDH>0.6

3. pleural fluid LDH more than two-thirds normal upper limit for serum

If a patient has an exudative pleural effusion, the following tests on the pleural fluid should be obtained: description of the fluid, glucose level, amylase level, differential cell count, microbiologic studies, and cytology.

Effusion due to Heart Failure The most common cause of pleural effusion is left

ventricular failure. The effusion occurs because the increased amounts of fluid in the lung interstitial spaces exit in part across the visceral pleura. This overwhelms the capacity of the lymphatics in the parietal pleura to remove fluid. A diagnostic thoracentesis should be performed if the effusions are not bilateral and comparable in size, if the patient is febrile, or if the patient has pleuritic chest pain to verify that the patient has a transudative effusion. Otherwise the patient is best treated with diuretics. If the effusion persists despite diuretic therapy, a diagnostic thoracentesis should be performed. Diuretic therapy for a few days does not significantly change the biochemical characteristics of the pleural fluid.

Hepatic Hydrothorax Pleural effusions occur in approximately 5% of patients with cirrhosis and ascites. The predominant mechanism is the direct movement of peritoneal fluid through small holes in the diaphragm into the pleural space. The effusion is usually right-sided and frequently is large enough to produce severe dyspnea. If medical management does not control the ascites and the effusion, the best treatment is a liver transplant. If the patient is not a candidate for this, the best alternative is insertion of a transjugular intrahepatic portal systemic shunt.

Parapneumonic Effusion Parapneumonic effusions are associated with bacterial pneumonia, lung abscess, or bronchiectasis and are probably the most common exudative pleural effusion in the United States. A *complicated parapneumonic effusion* requires tube thoracostomy for its resolution. *Empyema* refers to a grossly purulent effusion.

Patients with aerobic bacterial pneumonia and pleural effusion present with an acute febrile illness consisting of chest pain, sputum production, and leukocytosis. Patients with anaerobic infections present with a subacute illness with weight loss, a brisk leukocytosis, mild anemia, and a history of some factor that predisposes them to aspiration.

The possibility of a parapneumonic effusion should be considered whenever a patient with a bacterial pneumonia is initially evaluated. The presence of free pleural fluid can be demonstrated with a lateral decubitus radiograph. If the free fluid separates the lung from the chest wall by more than 10 mm on the decubitus radiograph, a therapeutic thoracentesis should be performed. Factors indicating the likely need for a procedure more invasive than a thoracentesis (in increasing order of importance) include:

1. loculated pleural fluid

2. pleural fluid pH below 7.20

3. pleural fluid glucose less than 60 mg/dL

4. positive Gram stain or culture of the pleural fluid

5. the presence of gross pus in the pleural space

If the fluid recurs after the initial therapeutic thoracentesis, a repeat thoracentesis should be performed if any of the above characteristics are present. If the fluid recurs a second

time, tube thoracostomy should be performed if any of the poor prognostic factors are present. If the fluid cannot be completely removed with the therapeutic thoracentesis, consideration should be given to inserting a chest tube and instilling a thrombolytic (streptokinase, 250,000 units or urokinase, 100,000 units) or performing thoracoscopy with the breakdown of adhesions. Decortication should be considered when the above are ineffective.

Effusion Secondary to Malignancy Malignant pleural effusions secondary to metastatic disease are the second most common type of exudative pleural effusion. The three tumors that cause approximately 75% of all malignant pleural effusions are lung carcinoma, breast carcinoma, and lymphoma. Most patients complain of dyspnea, which is frequently out of proportion to the size of the effusion. The pleural fluid is an exudate, and its glucose level may be reduced if the tumor burden in the pleural space is high.

The diagnosis is usually made via cytology of the pleural fluid. If the initial cytologic examination is negative, then thoracoscopy is the best next procedure if malignancy is strongly suspected. At the time of thoracoscopy, talc or some similar agent should be instilled into the pleural space to effect a pleurodesis. If thoracoscopy is unavailable, then needle biopsy of the pleura should be performed.

Patients with a malignant pleural effusion are treated symptomatically for the most part, since the presence of the effusion indicates disseminated disease and most malignancies associated with pleural effusion are not curable with chemotherapy. The only symptom that can be attributed to the effusion itself is dyspnea. If the patient's lifestyle is compromised by dyspnea, and if the dyspnea is relieved with a therapeutic thoracentesis, then one of the following procedures should be performed: (1) tube thoracostomy with the instillation of a sclerosing agent such as talc, 5 g in a slurry, or doxycycline, 500 mg; (2) outpatient insertion of a small indwelling catheter; or (3) thoracoscopy with pleural abrasion or the insufflation of talc.

Mesothelioma Malignant mesotheliomas are primary tumors that arise from the mesothelial cells that line the pleural cavities. Most are related to asbestos exposure. Patients with mesothelioma present with chest pain and shortness of breath. The chest radiograph reveals a pleural effusion, generalized pleural thickening, and a shrunken hemithorax. Thoracoscopy or open pleural biopsy is usually necessary to establish the diagnosis. Various treatment modalities, including radical surgery, chemotherapy, and radiation therapy, have been tried, but none has been proven to be more effective than symptomatic therapy. It is recommended that chest pain be treated with opiates and that shortness of breath be treated with oxygen and/or opiates.

Effusion Secondary to Pulmonary Embolization The diagnosis most commonly overlooked in the differential diagnosis of a patient with an undiagnosed pleural effusion is pulmonary embolism. Dyspnea is the most common symptom. The pleural fluid can be either transudative or exudative. The diagnosis is suggested by spiral CT scans, perfusion lung scanning and/or pulmonary arteriography (Chap. 261). Treatment of the patient with a pleural effusion secondary to pulmonary embolism is the same as for any patient with pulmonary emboli. If the pleural effusion increases in size after anticoagulation, the patient probably has recurrent emboli or another complication such as a hemothorax or a pleural infection.

Tuberculous Pleuritis (See also Chap. 169) In many parts of the world, the most common cause of an exudative pleural effusion is tuberculosis, but this is relatively uncommon in the United States. Tuberculous pleural effusions are thought to be due primarily to a hypersensitivity reaction to tuberculous protein in the pleural space. Patients with tuberculous pleuritis present with fever, weight loss, dyspnea, and/or pleuritic chest pain. The pleural fluid is an exudate with predominantly small lymphocytes. The diagnosis is established by demonstrating high levels of TB markers in the pleural fluid (adenosine deaminase > 45 IU/L, gamma interferon> 140 pg/mL, or positive PCR for tuberculous DNA). Alternatively, the diagnosis can be established by culture of the pleural fluid, needle biopsy of the pleura, or thoracoscopy. The recommended treatment of pleural and pulmonary tuberculosis is identical (Chap. 169).

Effusion Secondary to Viral Infection Viral infections are probably responsible for a sizable percentage of undiagnosed exudative pleural effusions. In many series, no diagnosis is established for approximately 20% of exudative effusions, and these effusions resolve spontaneously with no long-term residua. The importance of these effusions is that one should not be too aggressive in trying to establish a diagnosis for the undiagnosed effusion, particularly if the patient is improving clinically.

AIDS Pleural effusions are uncommon in such patients. The most common cause is Kaposi's sarcoma, followed by parapneumonic effusion. Other common causes are tuberculosis, cryptococcosis, and lymphoma. Pleural effusions are very uncommon with *Pneumocystis carinii* infection.

Chylothorax A chylothorax occurs when the thoracic duct is disrupted and chyle accumulates in the pleural space. The most common cause of chylothorax is trauma, but it also may result from tumors in the mediastinum. Patients with chylothorax present with dyspnea, and a large pleural effusion is present on the chest radiograph. Thoracentesis reveals milky fluid, and biochemical analysis reveals a triglyceride level that exceeds 110 mg/dL. Patients with chylothorax and no obvious trauma should have a lymphangiogram and a mediastinal computed tomographic (CT) scan to assess the mediastinum for lymph nodes. The treatment of choice for most chylothoraces is implantation of a pleuroperitoneal shunt. Patients with chylothoraces should not undergo prolonged tube thoracostomy with chest tube drainage because this will lead to malnutrition and immunologic incompetence.

Hemothorax When a diagnostic thoracentesis reveals bloody pleural fluid, a hematocrit should be obtained on the pleural fluid. If the hematocrit is>50% that of the peripheral blood, the patient has a hemothorax. Most hemothoraces are the result of trauma; other causes include rupture of a blood vessel or tumor. Most patients with hemothorax should be treated with tube thoracostomy, which allows continuous quantification of bleeding. If the bleeding emanates from a laceration of the pleura, apposition of the two pleural surfaces is likely to stop the bleeding. If the pleural hemorrhage exceeds 200 mL/h, consideration should be given to thoracotomy.

Miscellaneous Causes of Pleural Effusion There are many other causes of pleural effusion (Table 262-1). Key features of some of these conditions are as follows: If the pleural fluid amylase level is elevated, the diagnosis of esophageal rupture or pancreatic

disease is likely. If the patient is febrile, has predominantly polymorphonuclear cells in the pleural fluid, and has no pulmonary parenchymal abnormalities, an intraabdominal abscess should be considered. The diagnosis of an asbestos pleural effusion is one of exclusion. Benign ovarian tumors can produce ascites and a pleural effusion (Meigs' syndrome), as can the ovarian hyperstimulation syndrome. Several drugs can cause pleural effusion; the associated fluid is usually eosinophilic. Pleural effusions commonly occur following coronary artery bypass surgery. Effusions occurring within the first weeks are typically left-sided and bloody, with large numbers of eosinophils, and respond to one or two therapeutic thoracenteses. Effusions occurring after the first few weeks are typically left-sided and clear yellow, with predominantly small lymphocytes, and tend to recur. Other medical manipulations that induce pleural effusions include abdominal surgery, endoscopic variceal sclerotherapy, radiation therapy, liver or lung transplantation, or the intravascular insertion of central lines.

PNEUMOTHORAX

Pneumothorax is the presence of gas in the pleural space. A *spontaneous pneumothorax* is one that occurs without antecedent trauma to the thorax. A *primary spontaneous pneumothorax* occurs in the absence of underlying lung disease, while a *secondary spontaneous pneumothorax* occurs in its presence. A *traumatic pneumothorax* results from penetrating or nonpenetrating chest injuries. A *tension pneumothorax* is a pneumothorax in which the pressure in the pleural space is positive throughout the respiratory cycle.

Primary Spontaneous Pneumothorax Primary spontaneous pneumothoraces are usually due to rupture of apical pleural blebs, small cystic spaces that lie within or immediately under the visceral pleura. Primary spontaneous pneumothoraces occur almost exclusively in smokers, which suggests that these patients have subclinical lung disease. Approximately one-half of patients with an initial primary spontaneous pneumothorax will have a recurrence. The initial recommended treatment for primary spontaneous pneumothorax is simple aspiration. If the lung does not expand with aspiration, or if the patient has a recurrent pneumothorax, thoracoscopy with stapling of blebs and pleural abrasion is indicated. Thoracoscopy or thoracotomy with pleural abrasion is almost 100% successful in preventing recurrences.

Secondary Spontaneous Pneumothorax Most secondary spontaneous pneumothoraces are due to chronic obstructive pulmonary disease, but pneumothoraces have been reported with virtually every lung disease. Pneumothorax in patients with lung disease is more life-threatening than it is in normal individuals because of the lack of pulmonary reserve in these patients. Nearly all patients with secondary spontaneous pneumothorax should be treated with tube thoracostomy and the instillation of a sclerosing agent such as doxycycline or talc. Patients with secondary spontaneous pneumothoraces who have a persistent air leak, an unexpanded lung after 3 days of tube thoracostomy, or a recurrent pneumothorax should be subjected to thoracoscopy with bleb resection and pleural abrasion.

Traumatic Pneumothorax Traumatic pneumothoraces can result from both penetrating and nonpenetrating chest trauma. Traumatic pneumothoraces should be treated with tube thoracostomy unless they are very small. If a hemopneumothorax is present, one

chest tube should be placed in the superior part of the hemithorax to evacuate the air, and another should be placed in the inferior part of the hemithorax to remove the blood. Iatrogenic pneumothorax is a type of traumatic pneumothorax which is becoming more common. The leading causes are transthoracic needle aspiration, thoracentesis, and the insertion of central intravenous catheters. The treatment differs according to the degree of distress and can be observation, supplemental oxygen, aspiration, or tube thoracostomy.

Tension Pneumothorax This condition usually occurs during mechanical ventilation or resuscitative efforts. The positive pleural pressure is life-threatening both because ventilation is severely compromised and because the positive pressure is transmitted to the mediastinum, which results in decreased venous return to the heart and reduced cardiac output.

Difficulty in ventilation during resuscitation or high peak inspiratory pressures during mechanical ventilation strongly suggest the diagnosis. The diagnosis is made by the finding of an enlarged hemithorax with no breath sounds and shift of the mediastinum to the contralateral side. Tension pneumothorax must be treated as a medical emergency. If the tension in the pleural space is not relieved, the patient is likely to die from inadequate cardiac output or marked hypoxemia. A large-bore needle should be inserted into the pleural space through the second anterior intercostal space. If large amounts of gas escape from the needle after insertion, the diagnosis is confirmed. The needle should be left in place until a thoracostomy tube can be inserted.

DISORDERS OF THE MEDIASTINUM

The mediastinum is the region between the pleural sacs. It is separated into three compartments. The *anterior mediastinum* extends from the sternum anteriorly to the pericardium and brachiocephalic vessels posteriorly. It contains the thymus gland; the anterior mediastinal lymph nodes; and the internal mammary arteries and veins. The *middle mediastinum* lies between the anterior and posterior mediastina and contains the heart; the ascending and transverse arches of the aorta; the venae cavae; the brachiocephalic arteries and veins; the phrenic nerves; the trachea, main bronchi, and their contiguous lymph nodes; and the pulmonary arteries and veins. The *posterior mediastinum* is bounded by the pericardium and trachea anteriorly and the vertebral column posteriorly. It contains the descending thoracic aorta; esophagus; thoracic duct; azygos and hemiazygos veins; and the posterior group of mediastinal lymph nodes.

MEDIASTINAL MASSES

The first step in evaluating a mediastinal mass lesion is to place it in one of the three mediastinal compartments, since each has different characteristic lesions. The most common lesions in the anterior mediastinum are thymomas, lymphomas, teratomatous neoplasms, and thyroid masses. The most common masses in the middle mediastinum are vascular masses, lymph node enlargement from metastases or granulomatous disease, and pleuropericardial and bronchogenic cysts. In the posterior mediastinum, neurogenic tumors, meningoceles, meningomyeloceles, gastroenteric cysts, and esophageal diverticula are commonly found.

CTscanning is the most valuable imaging technique for evaluating mediastinal masses and is the only imaging technique that should be done in most instances. Barium studies of the gastrointestinal tract are indicated in many patients with posterior mediastinal lesions, since hernias, diverticula, and achalasia are readily diagnosed in this manner. An ^{131}I nuclear medicine scan can efficiently establish the diagnosis of intrathoracic goiter.

A definite diagnosis can be obtained with mediastinoscopy or anterior mediastinotomy in many patients with masses in the anterior or middle mediastinal compartments. A diagnosis can be established without thoracotomy via percutaneous fine-needle aspiration biopsy of mediastinal masses in any of the mediastinal compartments. In many cases the diagnosis can be established and the mediastinal mass removed with video-assisted thoracoscopy.

ACUTE MEDIASTINITIS

Most cases of acute mediastinitis are either due to esophageal perforation or occur after median sternotomy for cardiac surgery. Patients with esophageal rupture are acutely ill with chest pain and dyspnea due to the mediastinal infection. The esophageal rupture can occur spontaneously or as a complication of esophagoscopy or the insertion of a Blakemore tube. Appropriate treatment is exploration of the mediastinum with primary repair of the esophageal tear and drainage of the pleural space and the mediastinum.

The incidence of mediastinitis following median sternotomy is 0.4 to 5.0%. Patients most commonly present with wound drainage. Other presentations include sepsis or a widened mediastinum. The diagnosis is usually established with mediastinal needle aspiration. Treatment includes immediate drainage, debridement, and parenteral antibiotic therapy, but the mortality still exceeds 20%.

CHRONIC MEDIASTINITIS

The spectrum of chronic mediastinitis ranges from granulomatous inflammation of the lymph nodes in the mediastinum to fibrosing mediastinitis. Most cases are due to tuberculosis or histoplasmosis, but sarcoidosis, silicosis, and other fungal diseases are at times causative. Patients with granulomatous mediastinitis are usually asymptomatic. Those with fibrosing mediastinitis usually have signs of compression of some mediastinal structure such as the superior vena cava or large airways, phrenic or recurrent laryngeal nerve paralysis, or obstruction of the pulmonary artery or proximal pulmonary veins. Other than antituberculous therapy for tuberculous mediastinitis, no medical or surgical therapy has been demonstrated to be effective for mediastinal fibrosis.

PNEUMOMEDIASTINUM

In this condition, there is gas in the interstices of the mediastinum. The three main causes are: (1) alveolar rupture with dissection of air into the mediastinum; (2) perforation or rupture of the esophagus, trachea, or main bronchi; and (3) dissection of air from the neck or the abdomen into the mediastinum. Typically, there is severe substernal chest pain with or without radiation into the neck and arms. The physical

examination usually reveals subcutaneous emphysema in the suprasternal notch and *Hamman's sign*, which is a crunching or clicking noise synchronous with the heartbeat and best heard in the left lateral decubitus position. The diagnosis is confirmed with the chest radiograph. Usually no treatment is required, but the mediastinal air will be absorbed faster if the patient inspires high concentrations of oxygen. If mediastinal structures are compressed, the compression can be relieved with needle aspiration.

DISORDERS OF THE DIAPHRAGM

DIAPHRAGMATIC PARALYSIS

The presence of bilateral diaphragmatic paralysis almost always causes severe morbidity in adults. The most common causes include high spinal cord injury, thoracic trauma (including cardiac surgery), multiple sclerosis, anterior horn disease, and muscular dystrophy. Most patients with severe diaphragmatic weakness present with hypercapnic respiratory failure, frequently complicated by cor pulmonale and right ventricular failure, atelectasis, and pneumonia.

The degree of diaphragmatic weakness is best quantitated by measuring transdiaphragmatic pressures. The treatment of choice is assisted ventilation for all or part of each day. This is best accomplished without tracheostomy using nasal intermittent positive airway pressure. If the nerve to the diaphragm is intact, diaphragmatic pacing may be a viable alternative. If the paralysis occurs during open heart surgery, recovery frequently occurs, but it may take 6 months or more.

Unilateral paralysis of the diaphragm is much more common than is bilateral paralysis. The most common cause is nerve invasion from malignancy, usually a bronchogenic carcinoma. If the patient does not have malignancy, then usually no cause for the paralysis is found. The diagnosis is suggested by finding an elevated hemidiaphragm on the chest roentgenogram. Confirmation is best established with the "sniff test." When a patient is observed with fluoroscopy while sniffing, the paralyzed diaphragm will move paradoxically upward due to the negative intrathoracic pressure. Patients with a unilateral paralyzed diaphragm are usually asymptomatic. Their vital capacity and total lung capacity are each reduced about 25%. If a patient has a mediastinal mass in conjunction with the diaphragmatic paralysis, further workup should be done. However, if the patient is asymptomatic with a normal chest radiograph, no invasive procedures are warranted.

(Bibliography omitted in Palm version)

Back to Table of Contents

263. DISORDERS OF VENTILATION - *Eliot A. Phillipson*

HYPOVENTILATION

DEFINITION AND ETIOLOGY

Alveolar hypoventilation exists by definition when arterial P_{CO_2}(Pa_{CO_2}) increases above the normal range of 37 to 43 mmHg, but in clinically important hypoventilation syndromes Pa_{CO_2} is generally in the range of 50 to 80 mmHg. Hypoventilation disorders can be acute or chronic. The acute disorders, which represent life-threatening emergencies, are discussed in Chap. 265; this **chapter** deals with chronic hypoventilation syndromes.

Chronic hypoventilation can result from numerous disease entities (Table 263-1), but in all cases the underlying mechanism involves a defect in either the metabolic respiratory control system, the respiratory neuromuscular system, or the ventilatory apparatus. Disorders associated with impaired respiratory drive, defects in the respiratory neuromuscular system, some chest wall disorders such as obesity, and upper airway obstruction produce an increase in Pa_{CO_2}, despite normal lungs, because of a reduction in overall minute volume of ventilation and hence in alveolar ventilation. In contrast, most disorders of the chest wall and disorders of the lower airways and lungs may produce an increase in Pa_{CO_2}, despite a normal or even increased minute volume of ventilation, because of severe ventilation-perfusion mismatching that results in net alveolar hypoventilation.

Several hypoventilation syndromes involve combined disturbances in two elements of the respiratory system. For example, patients with chronic obstructive pulmonary disease may hypoventilate not simply because of impaired ventilatory mechanics but also because of a reduced central respiratory drive, which can be inherent or secondary to a coexisting metabolic alkalosis (related to diuretic and steroid therapy).

PHYSIOLOGIC AND CLINICAL FEATURES

Regardless of cause, the hallmark of all alveolar hypoventilation syndromes is an increase in alveolar P_{CO_2}(PA_{CO_2}) and therefore in Pa_{CO_2}(Fig. 263-1). The resulting respiratory acidosis eventually leads to a compensatory increase in plasma HCO_3^- concentration and a decrease in Cl^- concentration. The increase in PA_{CO_2} produces an obligatory decrease in PA_{O_2}, resulting in hypoxemia. If severe, the hypoxemia manifests clinically as cyanosis and can stimulate erythropoiesis and induce secondary polycythemia. The combination of chronic hypoxemia and hypercapnia may also induce pulmonary vasoconstriction, leading eventually to pulmonary hypertension, right ventricular hypertrophy, and congestive heart failure. The disturbances in arterial blood gases are typically magnified during sleep because of a further reduction in central respiratory drive. The resulting increased nocturnal hypercapnia may cause cerebral vasodilation leading to morning headache; sleep quality may also be severely impaired, resulting in morning fatigue, daytime somnolence, mental confusion, and intellectual impairment. Other clinical features associated with hypoventilation syndromes are related to the specific underlying disease (Table 263-1).

DIAGNOSIS

Investigation of the patient with chronic hypoventilation involves several laboratory tests that will usually localize the disorder to either the metabolic respiratory control system, the neuromuscular system, or the ventilatory apparatus (Fig. 263-2). Defects in the control system impair responses to chemical stimuli, including ventilatory, occlusion pressure, and diaphragmatic electromyographic (EMG) responses. During sleep, hypoventilation is usually more marked, and central apneas and hypopneas are common. However, because the behavioral respiratory control system (which is anatomically distinct from the metabolic control system), the neuromuscular system, and the ventilatory apparatus are intact, such patients can usually hyperventilate voluntarily, generate normal inspiratory and expiratory muscle pressures (PI_{max}, PE_{max}, respectively) against an occluded airway, generate normal lung volumes and flow rates on routine spirometry, and have normal respiratory system resistance and compliance and a normal alveolar-arterial P_{O2}[$(A\text{-}a)P_{O2}$] difference. Patients with defects in the respiratory neuromuscular system also have impaired responses to chemical stimuli but in addition are unable to hyperventilate voluntarily or to generate normal static respiratory muscle pressures, lung volumes, and flow rates. However, at least in the early stages of the disease, the resistance and compliance of the respiratory system and the alveolar-arterial oxygen difference are normal.

In contrast to patients with disorders of the respiratory control or neuromuscular systems, patients with disorders of the chest wall, lungs, and airways typically demonstrate abnormalities of respiratory system resistance and compliance and have a widened $(A\text{-}a)P_{O2}$. Because of the impaired mechanics of breathing, routine spirometric tests are abnormal, as is the ventilatory response to chemical stimuli. However, because the neuromuscular system is intact, tests that are independent of resistance and compliance are usually normal, including tests of respiratory muscle strength and of respiratory control that do not involve airflow.

TREATMENT

The management of chronic hypoventilation must be individualized to the patient's particular disorder, circumstances, and needs and should include measures directed toward the underlying disease. Coexistent metabolic alkalosis should be corrected, including elevations of HCO_3^- that are inappropriately high for the degree of chronic hypercapnia. Administration of supplemental oxygen is effective in attenuating hypoxemia, polycythemia, and pulmonary hypertension, but can aggravate CO_2 retention and the associated neurologic symptoms. For this reason, supplemental oxygen must be prescribed judiciously and the results monitored carefully. Pharmacologic agents that stimulate respiration (particularly progesterone) are of benefit in some patients, but generally, results are disappointing.

Most patients with chronic hypoventilation related to impairment of respiratory drive or neuromuscular disease eventually require mechanical ventilatory assistance for effective management. When hypoventilation is severe, treatment may be required on a 24-h basis, but in most patients ventilatory assistance only during sleep produces dramatic clinical improvement and lowering of daytime Pa_{CO2}. In patients with reduced respiratory drive but intact respiratory lower motor neurons, phrenic nerves, and

respiratory muscles, diaphragmatic pacing through an implanted phrenic electrode can be very effective. However, for patients with defects in the respiratory nerves and muscles, electrophrenic pacing is contraindicated. Such patients can usually be managed effectively with either intermittent negative-pressure ventilation in a cuirass or intermittent positive-pressure ventilation delivered through a tracheostomy or nose mask. For patients who require ventilatory assistance only during sleep, positive-pressure ventilation through a nose mask is the preferred method because it obviates a tracheostomy and avoids the problem of upper airway occlusion that can arise in a negative-pressure ventilator. Hypoventilation related to restrictive disorders of the chest wall (Table 263-1) can also be managed effectively with nocturnal intermittent positive-pressure ventilation through a nose mask or tracheostomy.

HYPOVENTILATION SYNDROMES

PRIMARY ALVEOLAR HYPOVENTILATION

Primary alveolar hypoventilation (PAH) is a disorder of unknown cause characterized by chronic hypercapnia and hypoxemia in the absence of identifiable neuromuscular disease or mechanical ventilatory impairment. The disorder is thought to arise from a defect in the metabolic respiratory control system, but few neuropathologic studies have been reported in such patients. Recent studies in animals suggest an important role for genetic factors in the pathogenesis of hypoventilation. Isolated PAH is relatively rare, and although it occurs in all age groups, the majority of reported cases have been in males aged 20 to 50 years. The disorder typically develops insidiously and often first comes to attention when severe respiratory depression follows administration of standard doses of sedatives or anesthetics. As the degree of hypoventilation increases, patients typically develop lethargy, fatigue, daytime somnolence, disturbed sleep, and morning headaches; eventually cyanosis, polycythemia, pulmonary hypertension, and congestive heart failure occur (Fig. 263-1). Despite severe arterial blood gas derangements, dyspnea is uncommon, presumably because of impaired chemoreception and ventilatory drive. If left untreated, PAH is usually progressive over a period of months to years and ultimately fatal.

The key diagnostic finding inPAH is a chronic respiratory acidosis in the absence of respiratory muscle weakness or impaired ventilatory mechanics (Fig. 263-2). Because patients can hyperventilate voluntarily and reduce Pa_{CO_2} to normal or even hypocapnic levels, hypercapnia may not be demonstrable in a single arterial blood sample, but the presence of an elevated plasma HCO_3-level should draw attention to the underlying chronic disturbance. Despite normal ventilatory mechanics and respiratory muscle strength, ventilatory responses to chemical stimuli are reduced or absent (Fig. 263-2), and breath-holding time may be markedly prolonged without any sensation of dyspnea.

Patients withPAHmaintain rhythmic respiration when awake, although the level of ventilation is below normal. However, during sleep, when breathing is critically dependent on the metabolic control system, there is typically a further deterioration in ventilation with frequent episodes of central hypopnea or apnea.

PAHmust be distinguished from other central hypoventilationsyndromes that are secondary to underlying neurologic disease of the brainstem or chemoreceptors (Table

263-1). This distinction requires a careful neurologic investigation for evidence of brainstem or autonomic disturbances. Unrecognized respiratory neuromuscular disorders, particularly those that produce diaphragmatic weakness, are often misdiagnosed as PAH. However, such disorders can usually be suspected on clinical grounds (see below) and can be confirmed by the finding of reduced voluntary hyperventilation, as well as PI_{max} and PE_{max}.

Some patients with PAH respond favorably to respiratory stimulant medications and to supplemental oxygen. However, the majority eventually require mechanical ventilatory assistance. Excellent long-term benefits can be achieved with diaphragmatic pacing by electrophrenic stimulation or with negative- or positive-pressure mechanical ventilation. The administration of such treatment only during sleep is sufficient in most patients.

RESPIRATORY NEUROMUSCULAR DISORDERS

Several primary disorders of the spinal cord, peripheral respiratory nerves, and respiratory muscles produce a chronic hypoventilation syndrome (Table 263-1). Hypoventilation usually develops gradually over a period of months to years and often first comes to attention when a relatively trivial increase in mechanical ventilatory load (such as mild airways obstruction) produces severe respiratory failure. In some of the disorders (such as motor neuron disease, myasthenia gravis, and muscular dystrophy), involvement of the respiratory nerves or muscles is usually a later feature of a more widespread disease. In other disorders, respiratory involvement can be an early or even isolated feature, and hence the underlying problem is often not suspected. Included in this category are the postpolio syndrome (a form of chronic respiratory insufficiency that develops 20 to 30 years following recovery from poliomyelitis), the myopathy associated with adult acid maltase deficiency, and idiopathic diaphragmatic paralysis.

Generally, respiratory neuromuscular disorders do not result in chronic hypoventilation unless there is significant weakness of the diaphragm. Distinguishing features of bilateral diaphragmatic weakness include orthopnea, paradoxical movement of the abdomen in the supine posture, and paradoxical diaphragmatic movement under fluoroscopy. However, the absence of these features does not exclude diaphragmatic weakness. Important laboratory features are a rapid deterioration of ventilation during a maximum voluntary ventilation maneuver and reduced PI_{max} and PE_{max} (Fig. 263-2). More sophisticated investigations reveal reduced or absent transdiaphragmatic pressures, calculated from simultaneous measurement of esophageal and gastric pressures; reduced diaphragmatic EMG responses (recorded from an esophageal electrode) to transcutaneous phrenic nerve stimulation; and marked hypopnea and arterial oxygen desaturation during rapid eye movement sleep, when there is normally a physiologic inhibition of all nondiaphragmatic respiratory muscles and breathing becomes critically dependent on diaphragmatic activity.

The management of chronic alveolar hypoventilation due to respiratory neuromuscular disease involves treatment of the underlying disorder, where feasible, and mechanical ventilatory assistance as described for the primary alveolar hypoventilation syndrome. However, electrophrenic diaphragmatic pacing is contraindicated in these disorders, except for high cervical spinal cord lesions in which the phrenic lower motor neurons and nerves are intact.

OBESITY-HYPOVENTILATION SYNDROME

Massive obesity represents a mechanical load to the respiratory system because the added weight on the rib cage and abdomen serves to reduce the compliance of the chest wall. As a result, the functional residual capacity (i.e., end-expiratory lung volume) is reduced, particularly in the recumbent posture. An important consequence of breathing at a low lung volume is that some airways, particularly those in the lung bases, may be closed throughout part or even all of each tidal breath, resulting in underventilation of the lung bases and widening of the (A-a)P_{O2}. Nevertheless, in the majority of obese individuals, central respiratory drive is increased sufficiently to maintain a normal Pa_{CO2}. However, a small proportion of obese patients develop chronic hypercapnia, hypoxemia, and eventually polycythemia, pulmonary hypertension, and right-sided heart failure. Recent studies in mice demonstrate that genetically obese mice lacking circulating leptin also develop chronic hypoventilation that can be reversed by leptin infusions. Those patients who also develop daytime somnolence have been designated as having the *Pickwickian syndrome* (Chap. 27). In many such patients, obstructive sleep apnea is a prominent feature, and even in those patients without sleep apnea, sleep-induced hypoventilation is an important element of the disorder and contributes to its progression. Most patients demonstrate a decrease in central respiratory drive, which may be inherent or acquired, and many have mild to moderate degrees of airflow obstruction, usually related to smoking. Based on these considerations, several therapeutic measures can be of considerable benefit, including weight loss, cessation of smoking, elimination of obstructive sleep apnea, and enhancement of respiratory drive by medications such as progesterone.

HYPERVENTILATION AND ITS SYNDROMES

DEFINITION AND ETIOLOGY

Alveolar hyperventilation exists when Pa_{CO2}decreases below the normal range of 37 to 43 mmHg. *Hyperventilation* is not synonymous with *hyperpnea*, which refers to an increased minute volume of ventilation without reference to Pa_{CO2}. Although hyperventilation is frequently associated with dyspnea, patients who are hyperventilating do not necessarily complain of shortness of breath; and conversely, patients with dyspnea need not be hyperventilating.

Numerous disease entities can be associated with alveolar hyperventilation (Table 263-2), but in all cases the underlying mechanism involves an increase in respiratory drive that is mediated through either the behavioral or the metabolic respiratory control systems (Fig. 263-3). Thus hypoxemia drives ventilation by stimulating the peripheral chemoreceptors, and several pulmonary disorders and congestive heart failure drive ventilation by stimulating afferent vagal receptors in the lungs and airways. Low cardiac output and hypotension stimulate the peripheral chemoreceptors and inhibit the baroreceptors, both of which increase ventilation. Metabolic acidosis, a potent respiratory stimulant, excites both the peripheral and central chemoreceptors and increases the sensitivity of the peripheral chemoreceptors to coexistent hypoxemia. Hepatic failure can also produce hyperventilation, presumably as a result of metabolic stimuli acting on the peripheral and central chemoreceptors.

Several neurologic and psychological disorders are thought to drive ventilation through the behavioral respiratory control system. Included in this category are psychogenic or anxiety hyperventilation and severe cerebrovascular insufficiency, which may interfere with the inhibitory influence normally exerted by cortical structures on the brainstem respiratory neurons. Rarely, disorders of the midbrain and hypothalamus induce hyperventilation, and it is conceivable that fever and sepsis also cause hyperventilation through effects on these structures. Several drugs cause hyperventilation by stimulating the central or peripheral chemoreceptors or by direct action on the brainstem respiratory neurons. Chronic hyperventilation is a normal feature of pregnancy and results from the effects of progesterone and other hormones acting on the respiratory neurons.

PHYSIOLOGIC AND CLINICAL FEATURES

Because hyperventilation is associated with increased respiratory drive, muscle effort, and minute volume of ventilation, the most frequent symptom associated with hyperventilation is dyspnea. However, there is considerable discrepancy between the degree of hyperventilation, as measured by P_{aCO_2}, and the degree of associated dyspnea. From a physiologic standpoint, hyperventilation is beneficial in patients who are hypoxemic, because the alveolar hypocapnia is associated with an increase in alveolar and arterial P_{O_2}. Conversely, hyperventilation can also be detrimental. In particular, the alkalemia associated with hypocapnia may produce neurologic symptoms, including dizziness, visual impairment, syncope, and seizure activity (secondary to cerebral vasoconstriction); parasthesia, carpopedal spasm, and tetany (secondary to decreased free serum calcium); and muscle weakness (secondary to hypophosphatemia). Severe alkalemia can also induce cardiac arrhythmias and evidence of myocardial ischemia. Patients with a primary respiratory alkalosis are also prone to periodic breathing and central sleep apnea (Chap. 264).

DIAGNOSIS

In most patients with a hyperventilation syndrome, the cause is readily apparent on the basis of history, physical examination, and knowledge of coexisting medical disorders (Table 263-2). In patients in whom the cause is not clinically apparent, investigation begins with arterial blood gas analysis, which establishes the presence of alveolar hyperventilation (decreased P_{aCO_2}) and its severity. Equally important is the arterial pH, which generally allows the disorder to be classified as either a primary respiratory alkalosis (elevated pH) or a primary metabolic acidosis (decreased pH). Also of importance is the P_{aO_2} and calculation of the $(A-a)P_{O_2}$, since a widened alveolar-arterial oxygen difference suggests a pulmonary disorder as the underlying cause. The finding of a reduced plasma HCO_3^- level establishes the chronic nature of the disorder and points toward an organic cause. Measurements of ventilation and arterial or transcutaneous P_{CO_2} during sleep are very useful in suspected psychogenic hyperventilation, since such patients do not maintain the hyperventilation during sleep.

The disorders that most frequently give rise to unexplained hyperventilation are pulmonary vascular disease (particularly chronic or recurrent thromboembolism) and psychogenic or anxiety hyperventilation. Hyperventilation due to pulmonary vascular disease is associated with exertional dyspnea, a widened $(A-a)P_{O_2}$ and maintenance of

hyperventilation during exercise. In contrast, patients with psychogenic hyperventilation typically complain of dyspnea at rest and not during mild exercise and of the need to sigh frequently. They are also more likely to complain of dizziness, sweating, palpitations, and paresthesia. During mild to moderate exercise, their hyperventilation tends to disappear and (A-a)P_{O2} is normal, but heart rate and cardiac output may be increased relative to metabolic rate.

TREATMENT

Alveolar hyperventilation is usually of relatively minor clinical consequence and therefore is generally managed by appropriate treatment of the underlying cause. In the few patients in whom alkalemia is thought to be inducing significant cerebral vasoconstriction, paresthesia, tetany, or cardiac disturbances, inhalation of a low concentration of CO_2 can be very beneficial. For patients with disabling psychogenic hyperventilation, careful explanation of the basis of their symptoms can be reassuring and is often sufficient. Others have benefited fromb-adrenergic antagonists or an exercise program. Specific treatment for anxiety may also be indicated.

(Bibliography omitted in Palm version)

Back to Table of Contents

264. SLEEP APNEA - *Eliot A. Phillipson*

DEFINITION AND CLASSIFICATION

Sleep apnea is defined as an intermittent cessation of airflow at the nose and mouth during sleep. By convention, apneas of at least 10 s duration have been considered important, but in most patients the apneas are 20 to 30 s in duration and may be as long as 2 to 3 min. *Sleep apnea syndrome* refers to a clinical disorder that arises from recurrent apneas during sleep. The clinical importance of sleep apnea arises from the fact that it is one of the leading causes of excessive daytime sleepiness. Indeed, epidemiologic studies have established a prevalence of clinically important sleep apnea of at least 2% in middle-aged women and 4% in middle-aged men.

Sleep apneas can be central or obstructive in type. In central sleep apnea (CSA) the neural drive to all the respiratory muscles is transiently abolished. In contrast, in obstructive sleep apnea (OSA) airflow ceases despite continuing respiratory drive because of occlusion of the oropharyngeal airway.

OBSTRUCTIVE SLEEP APNEA

Pathogenesis The definitive event inOSA is occlusion of the upper airway usually at the level of the oropharynx. The resulting apnea leads to progressive asphyxia until there is a brief arousal from sleep, whereupon airway patency is restored and airflow resumes. The patient then returns to sleep, and the sequence of events is repeated, often up to 400 to 500 times per night, resulting in marked fragmentation of sleep.

The immediate factor leading to collapse of the upper airway inOSA is the generation of a critical subatmospheric pressure during inspiration that exceeds the ability of the airway dilator and abductor muscles to maintain airway stability. During wakefulness, upper airway muscle activity is greater than normal in patients with OSA, presumably to compensate for airway narrowing (see below) and a high upper airway resistance. Sleep plays a permissive but crucial role by reducing the activity of the muscles and their protective reflex response to subatmospheric airway pressures. Alcohol is frequently an important cofactor because of its selective depressant influence on the upper airway muscles and on the arousal response that terminates each apnea. In most patients the patency of the airway is also compromised structurally and therefore predisposed to occlusion. In a minority of patients the structural compromise is due to obvious anatomic disturbances, such as adenotonsillar hypertrophy, retrognathia, and macroglossia. However, in the majority of patients the structural defect is simply a subtle reduction in airway size that can often be appreciated clinically as "pharyngeal crowding" and that can usually be demonstrated by imaging and acoustic reflection techniques. Obesity frequently contributes to the reduction in size of the upper airways, either by increasing fat deposition in the soft tissues of the pharynx or by compressing the pharynx by superficial fat masses in the neck. More sophisticated studies also demonstrate a high airway compliance -- i.e., the airway is "floppy" and therefore prone to collapse.

Pathophysiologic and Clinical Features The narrowing of the upper airways during sleep, which predisposes to OSA, inevitably results in snoring. In most patients, snoring

antedates the development of obstructive events by many years. However, the majority of snoring individuals do not have an OSA disorder, nor is there definitive evidence that snoring per se is associated with long-term health risks. Hence, in the absence of other symptoms, snoring alone does not warrant an investigation for OSA but does call for preventive counselling, particularly with regard to weight gain and alcohol consumption.

The recurrent episodes of nocturnal asphyxia and of arousal from sleep that characterizeOSA lead to a series of secondary physiologic events, which in turn give rise in some patients to the clinical complications of the syndrome (Fig. 264-1). The most common manifestations are neuropsychiatric and behavioral disturbances that are thought to arise from the fragmentation of sleep and loss of slow-wave sleep induced by the recurrent arousal responses. Nocturnal cerebral hypoxia may also play an important role. The most pervasive manifestation is excessive daytime sleepiness. Initially, daytime sleepiness manifests under passive conditions, such as reading or watching television; but as the disorder progresses, sleepiness encroaches into all daily activities and can become disabling and dangerous. Several studies have demonstrated two to seven times more motor vehicle accidents in patients with OSA compared with other drivers. Other related symptoms include intellectual impairment, memory loss, and personality disturbances.

The other major manifestations ofOSA are cardiorespiratory in nature and are thought to arise from the recurrent episodes of nocturnal asphyxia and of negative intrathoracic pressure, which increases left ventricular afterload (Fig. 264-1). Many patients demonstrate a cyclical slowing of the heart during the apneas to 30 to 50 beats per minute, followed by a tachycardia of 90 to 120 beats per minute during the ventilatory phase. A small number of patients develop severe bradycardia or dangerous tachyarrhythmias, leading to the notion that OSA may result in sudden death during sleep, but firm corroborative data are lacking. Unlike in healthy subjects, in patients with OSA systemic blood pressure fails to decrease during sleep. In fact, blood pressure typically rises abruptly at the termination of each obstructive event as a result of sympathetic nervous activation and reflex vasoconstriction. Furthermore, over 50% of patients with OSA have systemic hypertension. Several epidemiologic studies have implicated OSA as a risk factor for the development of systemic hypertension, and recent studies in an animal model demonstrate directly that OSA can cause sustained increases in daytime blood pressure. Emerging data also suggest that OSA can precipitate myocardial ischemia in patients with coronary artery disease and can adversely affect left ventricular function, both acutely and chronically, in patients with congestive heart failure. This complication is probably due to the combined effects of increased left ventricular afterload during each obstructive event, secondary to increased negative intrathoracic pressure (Fig. 264-1), recurrent nocturnal hypoxemia, and chronically elevated sympathoadrenal activity. Treatment of OSA in such patients often results in dramatic improvement in left ventricular function and in clinical cardiac status. Finally, up to 20% of patients with OSA develop mild pulmonary hypertension (in the absence of intrinsic lung disease), and a small proportion (<10%) develop pulmonary hypertension, right ventricular failure, polycythemia, and chronic hypercapnia and hypoxemia. All such patients have evidence of sustained daytime hypoxemia in addition to the nocturnal ventilatory disturbance, usually as a result of reduced ventilatory drive and/or diffuse airways obstruction.

Diagnosis AlthoughOSAoccurs at any age, and is more prevalent in women than was previously thought, the typical patient is a male aged 30 to 60 years who presents with a history of snoring, excessive daytime sleepiness, nocturnal choking or gasping, witnessed apneas during sleep, moderate obesity, and often mild to moderate hypertension. The definitive investigation for suspected OSA is polysomnography, a detailed overnight sleep study that includes recording of (1) electrographic variables (electroencephalogram, electrooculogram, and submental electromyogram) that permit the identification of sleep and its various stages, (2) ventilatory variables that permit the identification of apneas and their classification as central or obstructive, (3) arterial O_2saturation by ear or finger oximetry, and (4) heart rate. Continuous measurement of transcutaneous P_{CO2}(which reflects arterial P_{CO2}) can also be very useful, particularly in patients withCSA. The key diagnostic finding in OSA is episodes of airflow cessation at the nose and mouth despite evidence of continuing respiratory effort. By the time most patients come to clinical attention they have at least 10 to 15 obstructive events per hour of sleep. However, recent data suggest that a high upper airway resistance during sleep (manifested by snoring) that is accompanied by recurrent arousals from sleep, even in the absence of apneas and hypopneas, can result in a clinically important sleep-related syndrome. Therefore, the absence of outright apneas and hypopneas in a symptomatic patient may not definitely exclude a sleep-related respiratory disorder.

Because polysomnography is a time-consuming and expensive test, there is considerable interest in the role of simplified, unattended, ambulatory sleep monitoring for the investigation ofOSA that would allow the patient to be studied at home, rather than in the sleep laboratory. The most useful test in this context is the recording of arterial O_2saturation by oximetry. However, the reliability of overnight oximetry in the diagnosis of OSA is dependent on the pretest probability of the disorder. In patients with a high pretest probability (based on a history of daytime sleepiness, habitual snoring, nocturnal choking or gasping, and witnessed apneas during sleep), overnight oximetry can be used to *confirm* the diagnosis by demonstrating recurrent episodes of arterial O_2desaturation (at a rate of at least 10 to 15 events per hour). Such findings obviate the need for full polysomnography and allow initiation of treatment with nasal continuous positive airway pressure (CPAP) during sleep (see "Treatment"). However, negative results in a patient with a high clinical probability of OSA do not exclude the diagnosis but mandate that the patient proceed to polysomnography to investigate the cause of the daytime sleepiness. In contrast, when the pretest probability of OSA is low (such as the patient with only occasional snoring, few witnessed apneas, and no daytime sleepiness), the absence of arterial O_2desaturation can be used to *exclude* the diagnosis and thereby obviate the need for full polysomnography.

Studies suggest that overnight oximetry can obviate the need for polysomnography in about one-third of clinic patients referred for consideration ofOSA, either by *confirming* the diagnosis in patients with a *high* pretest probability of the disorder, or by *excluding* the diagnosis in patients with a *low* pretest probability. In the remaining two-thirds of patients with an intermediate pretest probability of OSA, overnight oximetry alone will not be definitive; hence such patients will require polysomnography.

TREATMENT

(Table 264-1) Several approaches to treatment ofOSA have been advocated, based on

an understanding of the mechanisms underlying the disorder. Mild to moderate OSA can often be managed effectively by modest weight reduction, avoidance of alcohol, improvement of nasal patency, and avoidance of sleeping in the supine posture. Intraoral appliances, designed to keep the mandible and tongue forward, are also effective in 55 to 80% of patients. The most widely used treatments in severe OSA are uvulopalatopharyngoplasty and nasalCPAPduring sleep. Uvulopalatopharyngoplasty is a surgical procedure designed to increase the pharyngeal lumen by resecting redundant soft tissue. When applied to unselected patients with OSA, it produces long-term cure in fewer than 50% but more discriminating selection of patients yields a higher rate of success. Other surgical approaches, including mandibular advancement and hyoid osteotomy have a more limited application but higher rate of success in selected patients. Nasal CPAP, which prevents upper airway occlusion by splinting the pharyngeal airway with a positive pressure delivered through a nose mask, is currently the most successful long-term approach to treatment, being well tolerated and effective in over 80% of patients, provided that they have received proper training. Patients who are unable to tolerate conventional nasal CPAP may respond to newer generation devices that provide more flexibility in adjusting the timing and levels of inspiratory and expiratory pressure cycles. For patients with ischemic heart disease or congestive heart failure who also have OSA, nasal CPAP is the only treatment that has been specifically tested and is considered the treatment of choice. Finally, for the few patients with severe OSA in whom all other treatment approaches have failed, tracheostomy can provide immediate relief, but in most centers is performed only very rarely.

CENTRAL SLEEP APNEA

Pathogenesis The definitive event inCSA is transient abolition of central drive to the ventilatory muscles. The resulting apnea leads to a primary sequence of events similar to those ofOSA(Fig. 264-1). Several underlying mechanisms can result in cessation of respiratory drive during sleep (Table 264-2). First are defects in the metabolic respiratory control system and respiratory neuromuscular apparatus. Such defects usually produce a chronic alveolar hypoventilation syndrome (in addition to CSA) that becomes more severe during sleep when the stimulatory effect of wakefulness on breathing is abolished. In contrast are CSA disorders that arise from transient instabilities in an otherwise intact respiratory control system. Common to all these disorders is a P_{CO_2}level during sleep that falls transiently below the critical P_{CO_2}required for respiratory rhythm generation. The most frequent instability of this type occurs at sleep onset, because the P_{CO_2}level of wakefulness is often lower than that required for rhythm generation in sleep; hence with loss of the stimulatory effect of wakefulness on breathing (referred to as the *waking neural drive*), an apnea develops at sleep onset until P_{CO_2}rises to the critical level (Fig. 264-2). However, if the central nervous system state fluctuates at sleep onset between "asleep" and "awake," a pattern of periodic breathing develops as respiration follows the changes in state. During each cycle, the waning phase of ventilation includes an hypopnea or outright central apnea (Cheyne-Stokes respiration). In most patients with CSA, the tendency to develop periodic breathing and central apneas during sleep is enhanced by some degree of chronic hyperventilation during wakefulness that drives the P_{CO_2}level below the threshold required for rhythm generation during sleep. Such hyperventilation is frequently idiopathic in nature. Hypoxia, whether due to high altitude or to underlying cardiorespiratory disease, also enhances the tendency to periodic breathing and CSA

for the same reasons. Periodic breathing and CSA are also common in patients with congestive heart failure. In such patients the decreases in P_{aCO_2} that trigger transient abolition of central respiratory drive are associated with higher left ventricular end-diastolic volume and filling pressure than in congestive heart failure patients without CSA. The hyperventilation probably results, therefore, from pulmonary congestion and stimulation of pulmonary vagal receptors.

Pathophysiologic and Clinical Features Many healthy individuals demonstrate a small number of central apneas during sleep, particularly at sleep onset and in rapid eye movement sleep. These apneas are not associated with any physiologic or clinical disturbances. In patients with clinically important CSA, the primary sequence of events that characterizes the disorder leads to prominent physiologic and clinical consequences (Fig. 264-1). In those patients whose CSA is a component of an alveolar hypoventilation syndrome, daytime hypercapnia and hypoxemia are usually evident, and the clinical picture is dominated by a history of recurrent respiratory failure, polycythemia, pulmonary hypertension, and right-sided heart failure. Complaints of sleeping poorly, morning headache, and daytime fatigue and sleepiness are also prominent. In contrast, in patients whose CSA results from an instability in respiratory drive, the clinical picture is dominated by features related to sleep disturbance, including recurrent nocturnal awakenings, morning fatigue, and daytime sleepiness. In patients with congestive heart failure, CSA can be an important (and frequently overlooked) cause of daytime sleepiness and fatigue. Recent studies also indicate that CSA can trigger sympathetic nervous activation in patients with heart failure and thereby exert a secondary deleterious effect on the underlying cardiac disorder.

Diagnosis Initially, many patients with CSA are suspected clinically of having OSA because of a history of snoring, sleep disturbance, and daytime sleepiness. However, obesity and hypertension are less prominent in CSA than in OSA. Definitive diagnosis of CSA requires a polysomnographic study, with the *key observation being recurrent apneas that are not accompanied by respiratory effort*. Measurements of transcutaneous P_{CO_2} are particularly useful in CSA. Those patients with a defect in respiratory control or neuromuscular function typically demonstrate an elevated P_{CO_2} that tends to increase progressively during the night, particularly during rapid eye movement sleep. In contrast, patients with instabilities in the respiratory control system typically demonstrate a mild degree of hypocapnia, which is an integral pathogenetic feature of their disorder (see above).

TREATMENT

The management of patients whose CSA is a component of an alveolar hypoventilation syndrome is essentially the same as management of the underlying hypoventilation disorder (Chap. 263). Management of patients whose CSA arises from an instability of respiratory drive is more problematic. Patients with hypoxemia usually respond favorably to nocturnal supplemental oxygen. Others have responded to acidification with acetazolamide, and recent reports indicate a good response to nasal CPAP (as for OSA). The mechanism by which CPAP abolishes central apneas probably involves a small increase in P_{aCO_2} as a result of the added expiratory mechanical load. In patients whose CSA is secondary to congestive heart failure, CPAP is particularly effective in improving sleep quality and daytime cardiac function. In fact, recent randomized trials have

demonstrated that CPAP has a beneficial effect on several surrogate markers of mortality in patients with congestive heart failure, including left ventricular ejection fraction, functional mitral regurgitation, and norepinephrine concentrations.

(Bibliography omitted in Palm version)

Back to Table of Contents

265. ACUTE RESPIRATORY DISTRESS SYNDROME - *Marc Moss*, *Roland H. Ingram, Jr.*

Lung injury in acute respiratory distress syndrome (ARDS) is characterized by increased permeability of the alveolar-capillary membrane, diffuse alveolar damage, and the accumulation of proteinaceous pulmonary edema. This clinical syndrome was first described in the archival literature by military physicians when respiratory failure occurred in battlefield casualties during World Wars I and II. However, it was not until the 1960s, when mechanical ventilation was used for patients with acute respiratory failure, that ARDS was first officially named. Initially the "A" in ARDS stood for "adult" to differentiate this syndrome from the infantile respiratory distress syndrome. With the more recent recognition that ARDS occurs in all age groups, the "A" now stands for "acute."

The diagnostic criteria used to defineARDShave evolved over the past three decades. Originally, most definitions required three general criteria: severe hypoxemia, decreased pulmonary compliance, and diffuse pulmonary infiltrates on chest radiograph. With the increasing utilization of pulmonary arterial catheters in the intensive care unit, ARDS was noted to be a "noncardiogenic" form of pulmonary edema (Chap. 32). Subsequently, some proposed definitions of ARDS required documentation of a normal pulmonary arterial occlusion pressure. However, due to the lack of an established definition and the recognition that ARDS is the severe form of a wide spectrum of lung injury, an American-European Consensus Conference proposed a new definition of ARDS that is now uniformly accepted (Table 265-1). Acute lung injury, which is a mild form of ARDS, was also defined and differs from ARDS based on less severe hypoxemia (Table 265-1).

CLINICAL CHARACTERISTICS

Many predisposing factors are associated with the development ofARDS, including conditions that injure the lung directly and those that produce damage through indirect mechanisms via the hematogenous delivery of inflammatory mediators (Table 265-2). The most common of these at-risk conditions are severe sepsis, major trauma, and aspiration of gastric contents. In general, 30 to 40% of individuals with at least one of these diagnoses will eventually develop ARDS. This incidence increases in patients with more than one at-risk condition. A history of chronic alcohol abuse is also associated with an increased risk of developing ARDS in critically ill patients with an at-risk diagnosis.

ARDSoccurs within 5 days of the initial at-risk diagnosis in the majority of patients, and over 50% will develop ARDS in the first 24 h. The earliest clinical sign is often an increase in the respiratory frequency, followed by dyspnea. There are no characteristic laboratory abnormalities for ARDS patients except those related to a specific underlying condition, such as leukocytosis in sepsis or an elevated serum amylase level in pancreatitis. Radiographically, the lung fields may be clear initially; diffuse bilateral interstitial or alveolar infiltrates occur as ARDS develops (Fig. 265-1). Though these radiographic changes appear homogeneous on chest radiograph, computed tomography demonstrates a heterogeneous pattern with a predominance of infiltrates in the dependent regions of the lung (Fig. 265-2).

PATHOPHYSIOLOGY

ARDSmay be the pulmonary manifestation of a systemic process and is the consequence of an overexpression of the normal inflammatory response. This inflammatory cascade has been divided into three overlapping phases -- initiation, amplification, and injury. During *initiation*, a precipitating event, such as sepsis, causes both immune and nonimmune cells to produce and release a variety of mediators and cytokines, such as tumor necrosis factor a and interleukin 1. Subsequently, during *amplification*, effector cells, such as neutrophils, are activated, recruited, and retained in specific target organs including the lung. Interleukin 8, which is produced by monocytes and other cell types, appears to play an important role in neutrophil activation. Once the effector cells have been sequestered in the lung, they then release reactive oxygen metabolites and proteases, causing cellular damage during the *injury phase*. This inflammatory cascade can occur systemically and therefore may alter the function of many organ systems -- a clinical entity called *multiple organ dysfunction syndrome*.

The pathophysiologic hallmark ofARDS is increased vascular permeability to proteins, so that even mild elevations of pulmonary capillary pressures (due to increased intravenous liquid administration and/or myocardial depression, which may occur in sepsis) greatly increase interstitial and alveolar edema. Alveolar damage is further exaggerated by the quantitative reduction is surfactant synthesis due to injury to type II pneumocytes as well as to further qualitative abnormalities in the size, composition, and metabolism of the remaining surfactant pool, leading to alveolar collapse. Although these atelectatic and liquid-filled regions of the lung contribute to a reduction in the compliance of the lung as a whole, significant regions of nondependent lung have relatively normal mechanical and gas-exchanging properties. However, the decreased overall pulmonary compliance requires large inspiratory pressures to be generated by the respiratory muscles, resulting in an increase in the work of breathing.

ThoughARDS is not routinely considered a disease of the airways, airway resistance may be increased due to bronchial wall edema and cytokine-mediated bronchospasm. Pulmonary vascular resistance and pulmonary arterial pressures may also be elevated as a result of increased pulmonary vascular smooth-muscle tone, perivascular edema, microvascular thrombosis, and the production of humoral factors such as leukotrienes and thromboxane A_2, which can directly cause vasoconstriction.

PATHOLOGY

During the initial exudative phase, covering the first few days after lung injury, the following occur: (1) epithelial cell injury represented by extensive necrosis of type I pneumocytes and a denuded basement membrane, (2) swelling of endothelial cells with the widening of intercellular junctions, (3) the formation of hyaline membranes composed of fibrin and other matrix proteins in alveolar ducts and airspaces, and (4) a neutrophilic inflammation. Fibrin thrombi may be seen in the alveolar capillaries and smaller pulmonary arteries. The second pathologic phase ofARDS is characterized by proliferation of a variety of cells and resolution of the neutrophilic inflammation. Cuboidal type II cells and squamous epithelium cover denuded alveolar basement membranes. Over the ensuing days to weeks, architectural restoration of lung tissue is usually

observed in survivors of ARDS. However, interstitial fibrosis and extensive restructuring of the lung parenchyma may occur with cystic and honeycomb changes in some ARDS patients, resulting in chronic pulmonary dysfunction or death.

TREATMENT

Currently there are no specific therapies that correct the underlying abnormalities in the permeability of the alveolar-capillary membrane or control the activated inflammatory response in patients withARDS. However, the use of physiologically targeted strategies of mechanical ventilation and intensive care unit management have led to a more favorable outcome for these critically ill patients.

Mechanical Ventilatory Support In the presence ofARDS, adequate oxygenation is not usually maintained when oxygen is supplied through noninvasive measures. Therefore, most ARDS patients require mechanical ventilation during their hospitalization. The primary goal of the ventilatory management in ARDS is to achieve ventilation and oxygenation that are adequate to support organ function. The major complications of mechanical ventilation are oxygen toxicity and barotrauma, which include not only pneumothorax, pneumomedistinum, and subcutaneous emphysema but also primary alveolar damage. As demonstrated on computed tomography images of the lungs in ARDS patients (Fig. 265-2), a large portion of the alveoli are atelectatic or liquid-filled. However, some nondependent regions of the lung remain radiographically unaffected, and due to their greater compliance they receive a greater proportion of the tidal volume. When large tidal volumes (10 to 12 mL/kg of ideal body weight) are forced into these smaller areas, damage may occur in epithelial and endothelial cells. The sequelae of this injury include alterations in lung liquid balance, increases in permeability, and severe alveolar damage. The deleterious effects of these large tidal volumes and subsequent high alveolar pressures has been termed *volutrauma*.

The currently recommended ventilatory strategies forARDSpatients focus on the limitation of airway pressures to a maximum inflation pressure that should not exceed 30 to 35 cmH_2O, rather than on strategies that attempt to achieve a normal Pa_{CO_2}. Because of the decreased overall lung compliance in ARDS patients, the use of low tidal volumes (~ 6 mL/kg of ideal body weight) is usually required. The subsequent decrease in minute ventilation may result in hypercapnia and respiratory acidosis. This ventilatory strategy, which emphasizes the limitation of transpulmonary pressures at the expense of hypercapnia, has been termed *permissive hypercapnia*.

After intubation, the inspired oxygen fraction (FI_{O_2}) is initially set at 1.0 and then decreased in steps to the lowest FI_{O_2}that will maintain an arterial oxygen tension (Pa_{O_2}) of approximately 60 mmHg. If Pa_{O_2}cannot be maintained at 60 mmHg by an FI_{O_2}£ 0.6, positive end-expiratory pressure (PEEP) may be added (Chap. 266). PEEP improves oxygenation by elevating mean alveolar pressure, thereby recruiting atelectatic alveoli and preventing end-expiratory airway and alveolar closure. In addition, PEEP may prevent alveolar damage by reducing the repetitive and cyclical reopening of closed alveoli during the respiratory cycle. Because PEEP may also overdistend uninvolved alveoli, it should be added cautiously, starting at 5 cmH_2O and increasing in increments of 3 to 5 cmH_2O to a maximum of 20 to 24 cmH_2O. Because airway pressure is transmitted to the pleural space, cardiac output may be adversely affected by the

addition of PEEP. In general, the optimal level of PEEP is the amount that achieves an acceptable arterial O_2saturation (³90%) with nontoxic FI_{O2}levels (£0.6) but without significantly compromising cardiac output. The comprehensive ventilatory strategy that combines low tidal volumes with adequate levels of PEEP has been termed a *lung-protective strategy*. ARDS patients ventilated with this technique have improved 28-day survival and require less time on mechanical ventilation when compared with ARDS patients treated with conventional ventilation using large tidal volumes achieving normal Pa_{CO2}levels.

Several other ventilatory strategies have been examined with the goal of improving oxygenation. However, none of these techniques has definitively been proven to be beneficial forARDSpatients. When turned from a supine to prone position, ARDS patients develop a more uniform distribution of pleural pressures, with an improvement in ventilation/perfusion matching and better postural drainage of secretions. Prone positioning may improve oxygenation in >75% of ARDS patients. However, the turning of these critically ill patients from the supine to the prone position is not without potential complications, such as unplanned extubation and removal of central venous catheters. The term *inverse ratio ventilation* is defined when the inspiratory (I) time exceeds the expiratory (E) time (i.e., > one-half of the respiratory cycle; I:E ratio > 1:1). This mode of ventilation is able to maintain a higher mean airway pressure, a major determinant of oxygenation, with lower peak airway pressures than conventional ventilation. However, due to the decrease in expiratory time, inverse ratio ventilation is potentially associated with dynamic hyperinflation and increases in end-expiratory pressure. Finally, partial liquid ventilation with perfluorocarbon, a radiopaque, inert, colorless liquid that carries a large quantity of O_2, and CO_2, has been studied in patients with severe ARDS. When perfluorocarbon is administered into the trachea of intubated patients, patients can be safely and adequately oxygenated and ventilated with routine mechanical ventilation.

Intravascular Volume Management Although pulmonary edema inARDSpatients is a consequence of increased permeability of the alveolar-capillary membrane, elevations in the intravascular hydrostatic pressure may also contribute to the accumulation of alveolar liquid and result in worsening oxygenation. Therefore, the optimal fluid management for patients with ARDS requires a balancing between liquid restriction, which may cause hypotension and decreased perfusion to vital organs, and liquid administration, which may increase oxygen requirements. Small decrements in the intravascular volume with diuretic use produce significant decreases in extravascular lung water. Caution must be exercised in reducing intravascular volume, since vigorous diuresis, especially in the setting ofPEEP, may reduce cardiac output and perfusion of critical organs. Ideally, the lowest intravascular hydrostatic pressure that also achieves an adequate cardiac output should be maintained. The placement of a pulmonary arterial catheter may be helpful in monitoring cardiac output and pulmonary arterial occlusion pressure (a measure of intravascular volume) in order to optimize the fluid management of patients with ARDS. However, the placement of a pulmonary arterial catheter and the clinical decisions based upon information derived from the catheter do not appear to improve and may actually worsen the outcome of general intensive care unit patients. Therefore the role of the pulmonary arterial catheter for ARDS patients is presently unclear.

Pharmacologic Therapies Due to their anti-inflammatory properties, glucocorticoids

have been used in patients withARDS, but when administered in high doses (30 mg/kg intravenously every 6 h for a total of four doses), they are not beneficial in the early course of the disease. In contrast, one small randomized study reported an improvement in mortality when glucocorticoids were given after 7 days of unresolving ARDS. In this study, active surveillance for infection was required before enrollment, and glucocorticoids were administered for up to 32 days. Future recommendations regarding the use of these drugs for ARDS patients will be based upon the results of an ongoing multicenter study.

Patients withARDShave both quantitative and qualitative abnormalities in surfactant, rendering surfactant-replacement therapy an attractive therapeutic modality. In one large randomized study of sepsis-induced ARDS, the administration of synthetic surfactant in an aerosolized form had no significant effect on outcome. Due to concerns with the efficacy of the delivery technique and the lack of essential surfactant-associated proteins in this particular replacement therapy, further studies of different surfactant preparations and modes of administration are presently ongoing.

When inhaled, nitric oxide vasodilates the pulmonary vasculature adjacent to well-ventilated alveoli, thereby improving ventilation-perfusion mismatching. Because of its subsequent inactivation by hemoglobin, nitric oxide produces a selective pulmonary vasodilation without systemic hemodynamic effects. Though inhaled nitric oxide appears to improve oxygenation initially, it is presently unknown whether this therapy will reduce mortality rates inARDSpatients.

PROGNOSIS

Since the initial descriptions ofARDS, mortality rates have ranged from 50 to 70%, although they may now be declining with optimal therapy. Mortality rates are higher in patients over 65 years of age, in those with an at-risk diagnosis of sepsis, and when associated with dysfunction of other organ systems. The cause of death for patients with ARDS has been traditionally divided into early causes (within 72 h) and late causes (after 3 days). Most early deaths are attributed to the original presenting illness or injury. Secondary infection and sepsis, persistent respiratory failure, and multiple-organ dysfunction are the most common causes of death in those ARDS patients who live at least 3 days.

In survivors ofARDS, abnormalities in pulmonary function normally improve considerably by 3 months and reach maximum levels of correction by 6 months after extubation. Although pulmonary function markedly recovers in many survivors, over 50% of these patients will continue to have abnormalities, including restrictive impairment or decreased diffusing capacity. Patients with severe ARDS, characterized by extreme hypoxemia and a longer duration of illness, usually have more pulmonary dysfunction than individuals with mild ARDS. Survivors of ARDS also have significant reductions in their quality of life, specifically in regard to physical functioning when compared to other previously critically ill patients.

(Bibliography omitted in Palm version)

Back to Table of Contents

266. MECHANICAL VENTILATORY SUPPORT - *Edward P. Ingenito*, *Jeffrey M. Drazen*

Ventilators are specially designed pumps that can support the ventilatory function of the respiratory system and improve oxygenation through application of high oxygen content gas and positive pressure. They are a mainstay of physiologic supportive care and are used to stabilize patients with respiratory failure as the underlying disease process is definitively treated.

INDICATIONS FOR MECHANICAL VENTILATION

Respiratory failure is the primary indication for initiation of mechanical ventilation. There are two basic types of respiratory failure.

Hypoxemic respiratory failure most commonly results from pulmonary conditions such as severe pneumonia, pulmonary edema, pulmonary hemorrhage, and respiratory distress syndrome causing ventilation-perfusion (/) mismatch and shunt. Hypoxemic respiratory failure is present when arterial O_2saturation (Sa_{O2})< 90% is observed despite an inspired O_2fraction (FI_{O2})> 0.6. The goal of ventilator treatment in this setting is to provide adequate Sa_{O2}through a combination of supplemental O_2 and specific patterns of ventilation that enhance oxygenation.

Hypercarbic respiratory failure results from disease states causing either a decrease in minute ventilation or an increase in physiologic dead space such that, despite adequate total minute ventilation, alveolar ventilation is inadequate to meet metabolic demands. Common clinical conditions associated with hypercarbic respiratory failure include neuromuscular diseases, such as myasthenia gravis, ascending polyradiculopathy, and myopathies, as well as diseases that cause respiratory muscle fatigue due to increased workload, such as asthma, chronic obstructive pulmonary disease, and restrictive lung disease. *Acute* hypercarbic respiratory failure is characterized by arterial P_{CO2}values of greater than 50 mmHg and an arterial pH above 7.30.

Mechanical ventilation generally should be instituted in acute hypercarbic respiratory failure. In contrast, the decision to institute mechanical ventilation when components of both acute and chronic hypercarbic respiratory failure are present depends on blood gas parameters and clinical evaluation. In particular, if a patient is not in respiratory distress and is not mentally impaired by CO_2accumulation, it is not mandatory to initiate mechanical ventilation while other forms of treatment are being administered. The goal of ventilator treatment in hypercarbic respiratory failure is to normalize arterial pH through changes in CO_2tensions. In patients with severe obstructive or restrictive lung disease, elevation in airway pressures may limit tidal volumes to the extent that normalization of pH is not possible, a situation known as *permissive hypercapnia*. Hypoxemic and hypercarbic respiratory failure may coexist in a given individual; in such cases, the indications for and goals of mechanical ventilation are similar to those in these two individual entities.

Accepted therapeutic applications of mechanical ventilation include controlled hyperventilation to reduce cerebral blood flow in patients with increased intracranial pressure or to improve pulmonary hemodynamics in patients with postoperative

pulmonary hypertension. Mechanical ventilation also has been used to reduce the work of breathing in patients with congestive heart failure, especially in the presence of myocardial ischemia. Ventilator support is also frequently used in conjunction with endotracheal intubation to prevent aspiration of gastric contents in otherwise unstable patients during gastric lavage for suspected drug overdose or during upper gastrointestinal endoscopy. In the critically ill patient, intubation and mechanical ventilation are indicated before essential diagnostic or therapeutic studies if it appears that respiratory failure may occur during these maneuvers.

PHYSIOLOGIC ASPECTS OF MECHANICAL VENTILATION

Most modern mechanical ventilators function by providing warmed and humidified gas to the airway opening in conformance with various specific volume, pressure, and time patterns. The ventilator serves as the energy source for inspiration, replacing the muscles of the diaphragm and chest wall. Expiration is passive, driven by the recoil of the lungs and chest wall; at the completion of inspiration, internal ventilator circuitry vents the airway to atmospheric pressure or a specified level of positive end-expiratory pressure (PEEP).

PEEPhelps maintain alveolar patency in the presence of destabilizing factors and therefore reverses hypoxemia and atelectasis by improving/)">/matching of ventilation and perfusion. PEEP levels between 0 and 10 cmH_2O are generally safe and effective; higher levels are recommended only in the management of significant refractory hypoxemia unresponsive to increments in FI_{O2} up to 0.6.

ESTABLISHING AND MAINTAINING AN AIRWAY

A cuffed endotracheal tube must be inserted to allow positive-pressure ventilators to deliver conditioned gas, at pressures above atmospheric pressure, to the lungs in a controlled fashion. If neuromuscular paralysis is to be induced during intubation, the use of agents whose mechanism of action includes depolarization at the neuromuscular junction, such as succinylcholine chloride, should be avoided in patients with renal failure, tumor lysis syndrome, crush injuries, medical conditions associated with elevated serum potassium levels, and muscular dystrophy syndromes. Opiates and benzodiazepines can have a deleterious effect on hemodynamics in patients with depressed cardiac function or low systemic vascular resistance and should be used cautiously in this setting. Morphine can promote histamine release from tissue mast cells and may worsen bronchospasm in patients with asthma; fentanyl, sufentanil, and alfentanil are acceptable alternatives to morphine. Ketamine may increase systemic arterial pressure as well as intracranial pressure and has been associated with dramatic hallucinatory responses; it should be used with caution in patients with hypertensive crisis, increased intracranial pressures, or a history of psychiatric disorders.

Patients who require ventilator support for extended periods of time may be candidates for tracheostomy. Although definitive guidelines for performing a tracheostomy in the ventilated patient have not been established, in current clinical practice patients who are anticipated to require ventilator therapy for more than 3 weeks should be considered for this procedure. While it does not clearly reduce the incidence of laryngeal injury or tracheal stenosis, tracheostomy has been associated with improved patient comfort and

enhanced ability to partake in rehabilitation-oriented activities.

VENTILATOR MODES

This setting specifies the manner in which ventilator breaths are triggered, cycled, and limited; commonly used modes of mechanical ventilation are given inTable 266-1. The *trigger*, either an inspiratory effort or a time-based signal, defines what the ventilator senses to initiate an assisted cycle. *Cycle* refers to the factors that determine the end of inspiration. For example, in volume-cycled ventilation, inspiration ends when a specific tidal volume is delivered to the patient. Other types of cycling include pressure cycling, time cycling, and flow cycling. *Limiting factors* are operator-specified values, such as airway pressure, that are monitored by transducers internal to the ventilator circuit throughout the respiratory cycle; if the specified values are exceeded, inspiratory flow is immediately stopped, and the ventilator circuit is vented to atmospheric pressure or the specifiedPEEP.

Assist Control Mode Ventilation (ACMV) An inspiratory cycle is initiated either by the patient's inspiratory effort or, if no patient effort is detected within a specified time window, by a timer signal within the ventilator. Every breath delivered consists of the operator-specified tidal volume. Ventilatory rate is determined either by the patient or by the operator-specified backup rate, whichever is of higher frequency (Fig. 266-1*A*). ACMV is the recommended mode for initiation of mechanical ventilation because it ensures a backup minute ventilation in the absence of an intact respiratory drive and allows for synchronization of the ventilator cycle with the patient's inspiratory effort.

Problems can arise whenACMV is used in patients with tachypnea due to nonrespiratory or nonmetabolic factors such as anxiety, pain, or airway irritation. Respiratory alkalemia may develop and trigger myoclonus or seizures. Dynamic hyperinflation (so-called auto-PEEP) may occur if the patient's respiratory mechanics are such that inadequate time is available for complete exhalation between inspiratory cycles. Auto-PEEP can limit venous return, decrease cardiac output, and increase airway pressures, predisposing to barotrauma. ACMV is not effective for weaning patients from mechanical ventilation because it provides full ventilator assistance on each patient-initiated breath.

Synchronized Intermittent Mandatory Ventilation (SIMV) The major difference between SIMV and ACMVis that in the former the patient is allowed to breathe spontaneously, i.e., without ventilator assist, between delivered ventilator breaths. However, mandatory breaths are delivered in synchrony with the patient's inspiratory efforts at a frequency determined by the operator. If the patient fails to initiate a breath, the ventilator delivers a fixed-tidal-volume breath and resets the internal timer for the next inspiratory cycle (Fig. 266-1*B*). SIMV differs from ACMV in that only the preset number of breaths is ventilator-assisted.

SIMVallows patients with an intact respiratory drive to exercise inspiratory muscles between assisted breaths. This characteristic makes SIMV a useful mode of ventilation for both supporting and weaning intubated patients. SIMV may be difficult to use in patients with tachypnea because they may attempt to exhale during the ventilator-programmed inspiratory cycle. When this occurs, the airway pressure may

exceed the inspiratory pressure limit, the ventilator-assisted breath will be aborted, and minute volume may drop below that programmed by the operator. In this setting, if the tachypnea is in response to respiratory or metabolic acidosis, a change toACMVwill increase minute ventilation and help normalize the pH while the underlying process is further evaluated.

Continuous Positive Airway Pressure (CPAP) This is not a true support-mode of ventilation, inasmuch as all ventilation occurs through the patient's spontaneous efforts. The ventilator provides fresh gas to the breathing circuit with each inspiration and charges the circuit to a constant, operator-specified pressure that can range from 0 to 20 cmH_2O (Fig. 266-1*C*). CPAP is used to assess extubation potential in patients who have been effectively weaned and are requiring little ventilator support and in patients with intact respiratory system function who require an endotracheal tube for airway protection.

Pressure-Control Ventilation (PCV) This form of ventilation is time triggered, time cycled, and pressure limited. During the inspiratory phase, a given pressure is imposed at the airway opening, and the pressure remains at this user-specified level throughout inspiration (Fig. 266-2*A*). Since inspiratory airway pressure is specified by the operator, tidal volume and inspiratory flow rate are *dependent* rather than *independent* variables and are not user specified. PCV is the preferred mode of ventilation for patients with documented barotrauma, because airway pressures can be limited, and for postoperative thoracic surgical patients, in whom the shear forces across a fresh suture line should be limited. When PCV is used, minute ventilation and tidal volume must be monitored; minute ventilation is altered through changes in rate or in the pressure-control value.

The major practical limitation ofPCV is patient-ventilator asynchrony related to its time-cycled and time-triggered characteristics. Because PCV requires that the patient passively accept ventilator breaths, most patients require heavy sedation to be maintained on this ventilatory mode, which may be hazardous in the hemodynamically unstable patient.

PCVwith the use of a prolonged inspiratory time is frequently applied to patients with severe hypoxemic respiratory failure. This approach, called inverse inspiratory-to-expiratory ratio ventilation (IRV), increases mean distending pressures without increasing peak airway pressures. It is thought to work in conjunction withPEEP to open collapsed alveoli and improve oxygenation. IRV may be associated with fewer deleterious effects than conventional volume-cycled ventilation, which requires higher peak airway pressures to achieve an equivalent reduction in shunt fraction.

Pressure-Support Ventilation (PSV) This form of ventilation is patient triggered, flow cycled, and pressure limited; it is specifically designed for use in the weaning process. During PSV, the inspiratory phase is terminated when inspiratory airflow falls below a certain level; in most ventilators this flow rate cannot be adjusted by the operator. When PSV is used, patients receive ventilator assist only when the ventilator detects an inspiratory effort (Fig. 266-2*B*). PSV also can be used in combination withSIMV to ensure volume-cycled backup for patients whose respiratory drive is depressed either spontaneously or as a result of various therapeutic maneuvers.

PSVis well tolerated by most patients who are being weaned; PSV parameters can be set to provide fully or nearly fully ventilatory support and can be withdrawn slowly over a period of days in a systematic fashion to gradually load the respiratory muscles.

Open Lung Ventilation (OLV) OLV is not a distinct mode of ventilation, but rather a strategy for applying either volume-cycled or pressure-control ventilation to patients with severe respiratory failure. In OLV, the primary objectives of ventilator support are maintenance of adequate oxygenation and avoidance of cyclic opening and closing of alveolar units by selecting a level ofPEEPthat allows the majority of units to remain inflated during tidal ventilation. Achievement of eucapnia and normal blood pH through adjustments in ventilator tidal volume and breathing frequency are of lower priority. Clinical and experimental observations indicate that high airway pressures and repeated opening and closing of alveoli can cause microstructural lung damage, propagation of lung injury through generation of inflammatory cytokines, and direct barotrauma. Current data suggest that a small tidal volume (i.e., 6 mL/kg) provides adequate ventilatory support with a lower incidence of adverse effects than more conventional tidal volumes of 10-15 mL/kg. These potential complications can have dire consequences in patients with respiratory failure. Alternatively, hypercapnia and consequent respiratory acidosis tend to be well tolerated physiologically, except in patients with significant hemodynamic compromise, ventricular dysfunction, cardiac dysrrhythmias, or increased intracranial pressure. OLV has been used most extensively in the management of patients with hypoxemic respiratory failure due to acute lung injury. Although few randomized clinical trials of OLV have been performed, available data suggest that OLV reduces the morality rate and improves gas exhange in patients with acute lung injury.

Prone Positioning during Mechanical Ventilation Patients with acute respiratory distress syndrome (ARDS) experience hypoxemia as a result of intrapulmonary shunt due to regional atelectasis. Recent studies in patients with ARDS have demonstrated that collapse occurs most extensively in the dependent regions of the lung. Increasing airway pressures to counterbalance the compressive effects of the surrounding lung in these collapsed regions improves gas exchange but may result in potentially dangerous peak airway pressures. Prone positioning, in both experimental and clinical studies, reduces shunt and improves oxygenation by causing regional improvements in transpulmonary distending pressures wtihout overexpanding already patent alveoli. In clinical practice, prone positioning has been used in conjunction with both volume-cycled and pressure-control ventilation with equivalent clinical effectiveness and appears to be a useful adjunct to conventional ventilator support in patients with severe hypoxemic respiratory failure.

Noninvasive Ventilation (NIV) Noninvasive ventilator support through a tight-fitting facemask or nasal mask, traditionally used for treatment of sleep apnea, has recently been used as primary ventilator support in patients with impending respiratory failure. Facemask and nasal devices for administering NIV therapy are most frequently combined withPSV or bi-level positive airway pressure ventilation, inasmuch as both of these modes are well tolerated by the conscious patient and optimize patient-ventilator synchrony. NIV has met with varying degrees of success when applied to patients with acute or chronic respiratory failure. The major limitation to its widespread application has been patient intolerance, because the tight-fitting mask required for NIV can cause

both physical and emotional discomfort in patients with dyspnea. In general, centers with experience using NIV have reported clinical success with minimal associated morbidity, whereas centers with less experience have reported more limited success. Aggressive medical therapy directed at the cause of impending respiratory failure, together with an experienced respiratory therapy and physician team, appear to be the keys to successful use of NIV in intensive care units.

Extracorporeal Membrane Oxygenation (ECMO) This nonconventional mode of ventilator support employs a large surface area membrane system connected in series with the patient's circulation to exchange CO_2 and O_2. The lung functions primarily as a passive conduit with gas exchange occurring by diffusion across the membrane. ECMO was first examined in 1970 as an alternative to positive-pressure ventilation in the management of patients withARDS. Initial studies failed to demonstrate an improvement in survival rates among patients treated with ECMO. Although several uncontrolled trials have since suggested that ECMO does improve outcome among patients with ARDS, a 1993 study comparing survival rates of patients with ARDS treated with ECMO and those treated with conventional ventilator therapy showed no difference in mortality rates, but the morbidity rates and hospital costs were increased among ECMO-treated patients. Presently, the use of ECMO in patients with ARDS is not recommended.

GUIDELINES FOR MANAGING THE VENTILATED PATIENT

Most patients who are started on ventilator support receiveACMV orSIMV, because these modes ensure user-specified backup minute ventilation in the event that the patient fails to initiate respiratory efforts. Once the intubated patient has been stabilized with respect to oxygenation, definitive therapy for the underlying process responsible for respiratory failure is formulated and initiated. Subsequent modifications in ventilator therapy must be provided in parallel with changes in the patient's clinical status. As improvement in respiratory function is noted, the first priorities are to reducePEEP and supplemental O_2. Once a patient can achieve adequate arterial saturation with an FIO_2£ 0.5 and 5 cmH_2O PEEP, attempts should be made to reduce the level of mechanical ventilatory support. Patients previously on full ventilator support should be switched to a ventilator mode that allows for weaning, such as SIMV,PSV, or SIMV combined with PSV. Ventilator therapy can then be gradually removed, as outlined in the section on weaning. Patients whose condition continues to deteriorate after ventilator support is initiated may require increased O_2, PEEP, and alternative modes of ventilation such asIRV.

GENERAL SUPPORT IN THE VENTILATED PATIENT

Patients who are started on mechanical ventilation usually require some form of sedation and analgesia to maintain an acceptable level of comfort. Often, this regimen consists of a combination of a benzodiazepine and opiate administered intravenously. Medications commonly used for this purpose include lorazepam, midazolam, diazepam, morphine, and fentanyl.

Immobilized patients in the intensive care unit on mechanical ventilator support are at increased risk for deep venous thrombosis; accepted practice consists of administering prophylaxis in the form of subcutaneous heparin and/or pneumatic compression boots.

Fractionated low molecular weight heparin has also been used for this purpose; it appears to be equally effective and is associated with a decreased incidence of heparin-associated thrombocytopenia.

Prophylaxis against diffuse gastrointestinal mucosal injury is indicated for patients who have suffered a neurologic insult or those with severe respiratory failure in association withARDS. Histamine receptor antagonists (H_2-receptor antagonists), antacids, and cytoprotective agents such as carafate have all been used for this purpose and appear to be effective. Recent data suggest that carafate use is associated with a reduction in the incidence of nosocomial pneumonias, since it does not cause changes in stomach pH and is less likely to permit colonization of the gastrointestinal tract by nosocomial organisms at pH levels near neutral.

Nutrition support by enteral feeding through either a nasogastric or an orogastric tube should be maintained in all intubated patients whenever possible. In those patients with a normal baseline nutritional state, support should be initiated within 7 days. In malnourished patients, nutrition support should be initiated within 72 h. Delayed gastric emptying is common in critically ill patients on sedative medications but often responds to promotility agents such as cisapride or metoclopramide. Parenteral nutrition is an alternative to enteral nutrition in patients with severe gastrointestinal pathology.

COMPLICATIONS OF MECHANICAL VENTILATION

Endotracheal intubation and positive-pressure mechanical ventilation have direct and indirect effects on several organ systems, including the lung and upper airways, the cardiovascular system, and the gastrointestinal system. Pulmonary complications include barotrauma, nosocomial pneumonia, oxygen toxicity, tracheal stenosis, and deconditioning of respiratory muscles. *Barotrauma*, which occurs when high pressures (i.e., > 50 cmH_2O) disrupt lung tissue, is clinically manifest by interstitial emphysema, pneumomediastinum, subcutaneous emphysema, or pneumothorax. Although the first three conditions may resolve simply through the reduction of airway pressures, clinically significant pneumothorax, as indicated by hypoxemia, decreased lung compliance, and hemodynamic compromise, requires tube thoracostomy.

Patients intubated for longer than 72 h are at high risk for *nosocomial pneumonia* as a result of aspiration from the upper airways through small leaks around the endotracheal tube cuff; the most common organisms responsible for this condition are enteric gram-negative rods, *Staphylococcus aureus*, and anaerobic bacteria. Because the endotracheal tube and upper airways of patients on mechanical ventilation are commonly colonized with bacteria, the diagnosis of nosocomial pneumonia requires "protected brush" bronchoscopic sampling of airway secretions coupled with quantitative microbiologic techniques to differentiate colonization from infection.

Oxygen toxicity is a potential complication when an FI_{O_2}[3] 0.6 is required for more than 72 h. The condition can be prevented in some cases through the use ofPEEP to allow for FI_{O_2}values to go below 0.6 while primary therapy for the underlying condition is instituted. Although O_2toxicity is thought to result from the effects of oxygen free radical on the lung interstitium, the therapeutic use of antioxidants such as superoxide dismutase, catalase, selenium, and vitamin E remains experimental.

Hypotension resulting from elevated intrathoracic pressures with decreased venous return is almost always responsive to intravascular volume repletion. In patients judged to have hypotension or respiratory failure on the basis of alveolar edema, hemodynamic monitoring with a pulmonary arterial catheter may be of value in optimizing O_2delivery via manipulation of intravascular volume and FI_{O2} andPEEPlevels.

Gastrointestinal effects of positive-pressure ventilation include *stress ulceration* and *mild to moderate cholestasis*. It is common practice to provide prophylaxis with H_2-receptor antagonists or sucralfate for stress-related ulcers. Mild cholestasis (i.e., total bilirubin values £4.0) attributable to the effects of increased intrathoracic pressures on portal vein pressures is common and generally self-limited. Cholestasis of a more severe degree should not be attributed to a positive-pressure ventilation response and is more likely due to a primary hepatic process.

WEANING FROM MECHANICAL VENTILATION

Removal of mechanical ventilator support requires that a number of criteria be met. Upper airway function must be intact for a patient to remain extubated but is difficult to assess in the intubated patient. Therefore, if a patient can breathe on his or her own through an endotracheal tube but develops stridor or recurrent aspiration once the tube is removed, upper airway dysfunction or an abnormal swallowing mechanism should be suspected and plans for achieving a stable airway developed. An intact cough during suctioning is a good indicator of a patient's ability to mobilize secretions. Respiratory drive and chest wall function are assessed by observation of respiratory rate, tidal volume, inspiratory pressure, and vital capacity. The weaning index, defined as the ratio of breathing frequency to tidal volume (breaths per minute per liter), is both sensitive and specific for predicting the likelihood of successful extubation. When this ratio is less than 105 with the patient breathing without mechanical assistance through an endotracheal tube, successful extubation is likely. An inspiratory pressure of more than -30 cmH_2O and a vital capacity of greater than 10 mL/kg are considered indicators of acceptable chest wall and diaphragm function. Alveolar ventilation is generally adequate when elimination of CO_2 is sufficient to maintain arterial pH in the range of 7.35 to 7.40, and anSa_{O2}> 90% can be achieved with an FI_{O2}< 0.5 and a$PEEP$£ 5 cmH_2O. Although many patients may not meet all criteria for weaning, the likelihood that a patient will tolerate extubation without difficulty increases as more criteria are met.

Many approaches to weaning patients from ventilator support have been advocated. T-piece andCPAPweaning are best tolerated by patients who have undergone mechanical ventilation for brief periods and require little respiratory muscle reconditioning, whereasSIMV andPSV are best for patients who have been intubated for extended periods and require gradual respiratory-muscle reconditioning.

T-piece weaning involves brief spontaneous breathing trials with supplemental O_2. These trials are usually initiated for 5 min/h followed by a 1-h interval of rest. T-piece trials are increased in 5- to 10-min increments until the patient can remain ventilator independent for periods of several hours. Extubation can then be attempted.CPAPweaning is similar to T-piece weaning except that trials of spontaneous breathing are conducted on the ventilator in CPAP mode.

Weaning by means ofSIMVinvolves gradually tapering the mandatory backup rate in increments of 2 to 4 breaths per minute while monitoring blood gas parameters and respiratory rates. Rates of greater than 25 breaths per minute on withdrawal of mandatory ventilator breaths generally indicate respiratory muscle fatigue and the need to combine periods of exercise with periods of rest. Exercise periods are gradually increased until a patient remains stable on SIMV at 4 breaths per minute or less without needing rest at higher SIMV rates. ACPAP or T-piece trial can then be attempted before planned extubation.

PSV, as described in detail above, is used primarily for weaning from mechanical ventilation. PSV is usually initiated at a level adequate for full ventilator support (PSV_{max}); i.e., PSV is set slightly below the peak inspiratory pressures required by the patient during volume-cycled ventilation. The level of pressure support is then gradually withdrawn in increments of 5 cmH_2O until a level is reached at which the respiratory rate increases to 25 breaths per minute. At this point, intermittent periods of higher-pressure support are alternated with periods of lower-pressure support to provide muscle reconditioning without causing diaphragmatic fatigue. Gradual withdrawal of PSV continues until the level of support is just adequate to overcome the resistance of the endotracheal tube (approximately 5 to 10 cmH_2O). Support can be discontinued and the patient extubated.

(Bibliography omitted in Palm version)

Back to Table of Contents

267. LUNG TRANSPLANTATION - *Janet R. Maurer*

Lung transplantation for end-stage lung disease has been a therapeutic option since the 1980s. Several transplant options are available for carefully selected patients: unilateral lung transplant, bilateral lung transplant, heart-lung transplant (Chap. 233), and living lobar transplant. The first successful type of lung transplant was heart-lung, which was performed for a variety of indications and in increasing numbers until 1989. Beginning in 1989, the numbers of both unilateral and bilateral lung transplants performed increased dramatically and, along with the increasing demand for heart donors, greatly reduced the number of donor organs available for heart-lung procedures. By the mid-1990s, unilateral and bilateral lung transplant numbers also had plateaued because of donor shortages and have remained relatively stable at approximately 1250 operations worldwide per year. Of these, 60% are unilateral lung transplants and 40% are bilateral. Heart-lung transplants have leveled off at between 100 and 150 per year. In its 1999 report, the Registry of the International Society for Heart and Lung Transplantation, in conjunction with the United Network for Organ Sharing, had recorded a cumulative total of 8997 isolated lung transplants and 2350 heart-lung transplants.

INDICATIONS

Emphysema, either smoking-induced or secondary to α_1-antitrypsin deficiency, has been the single largest indication for lung transplantation. This diagnosis accounts for about 55% of unilateral lung transplants, 29% of bilateral, and 6% of heart-lung transplants. Other major indications for unilateral lung transplants include idiopathic pulmonary fibrosis (21%) and primary pulmonary hypertension (5%). Patients with cystic fibrosis comprise the largest group of bilateral lung recipients, approximately 34% of the total, followed by patients with emphysema, patients with primary pulmonary hypertension (10%), and those with idiopathic pulmonary fibrosis (7.5%). The major diagnoses among heart-lung recipients are primary and secondary pulmonary hypertension (54%) and cystic fibrosis (16%). With the widespread use of unilateral and bilateral lung transplants, the indications for heart-lung transplant have become very circumscribed, so that now most candidates for this type of transplant have either concomitant left ventricular disease and end-stage lung disease or irreparable congenital heart disease with Eisenmenger's syndrome. Patients receiving living lobar donations have been either children or young adults, and most have suffered from cystic fibrosis. In most of these operations, a lower lobe is donated from each of two adults, who are often, but not always, related to the recipient. Living donation is performed in a limited number of lung transplant programs, and the donor morbidity rate has been acceptable.

RECIPIENT SELECTION

Because donor lungs are the scarcest of the common solid organs transplanted, patients with end-stage lung disease undergo extensive evaluation to select the best potential candidates. In 1998, this process was further standardized with the publication of the International Guidelines for the Selection of Lung Transplant Candidates. Approximate age limits of 65 years for unilateral lung, 60 years for bilateral lung, and 55 years for heart-lung transplants were set. Chronic medical conditions that can be adequately controlled and have not resulted in end-organ damage, e.g., systemic hypertension, are acceptable in lung transplant candidates. However, in a case where a

chronic illness is often associated with nonpulmonary organ damage, e.g., diabetes mellitus, a careful assessment of target organ function is necessary.

Absolute contraindications to lung transplantation include dysfunction of major organs (other than lung), infection with HIV, active malignancy within 2 years with the exception of basal cell and squamous cell skin cancer, hepatitis B antigen positivity, and hepatitis C with biopsy-proven histologic evidence of liver disease. Conditions that represent relative contraindications include symptomatic osteoporosis; severe musculoskeletal disease affecting the thorax; high-dose corticosteroid use; weight less than 70% or greater than 130% of ideal body weight; alcohol, cigarette, or narcotic abuse/addiction within 6 months before evaluation; psychosocial problems, including noncompliance, that cannot be adequately resolved through pharmacologic treatment or counseling; requirement for invasive ventilation; and colonization with fungi or atypical mycobacteria. Colonization is of particular concern when a unilateral lung transplant is being considered. Disease-specific guidelines (Table 267-1) are chosen to identify candidates who are within the transplant "window" -- that is, patients who are ill enough to fit within the category of "end-stage" and have progressive disease, yet are able to survive the pre-transplant waiting and perioperative time periods. In the past few years, increasing experience with large numbers of patients with end-stage disease has made it much easier to estimate life expectancies; however, patients with diagnoses of emphysema and Eisenmenger's type pulmonary hypertension remain problematic in this regard because posttransplant statistical analysis does not show a clear survival benefit for recipients within the first 2 years. In these types of patient, selection usually includes consideration of quality-of-life issues as well as survival rates.

SELECTION OF TRANSPLANT PROCEDURE

The only diseases that currently mandate a specific procedure are (1) irreparable congenital cardiac defects with Eisenmenger's syndrome (heart-lung transplant); (2) advanced lung disease with concomitant left ventricular dysfunction (heart-lung transplant); and (3) bronchiectatic lung disease, e.g., cystic fibrosis (bilateral lung transplant or bilobar living donor lung transplant). In essentially all other circumstances unilateral lung transplantation can be performed with acceptable early and midterm results. Bilateral lung transplantation, however, is often preferred if difficulty is anticipated in postoperative management, especially in patients with pulmonary hypertension; if significant bullous disease in present in emphysema; if a patient is very young; or if there are specific individual recipient considerations. As noted below, the long-term survival rates of bilateral lung recipients may be superior; nevertheless, transplant centers have generally chosen to maximize the donor organ resource by performing unilateral lung transplants whenever possible, rather than opting for potentially slightly increased survival periods.

PROGNOSIS

The 1- and 2-year survival rates for unilateral and bilateral lung transplant recipients are 67 and 62%, respectively. Longer term data show a divergence in survival rates by 5 years, with the half-life of bilateral transplant recipients (4.9 years) significantly longer than that of unilateral transplant recipients (3.6 years). Among unilateral graft recipients, patients with emphysema appear to have the best early survival rate (nearly 80% at one

year), and patients with idiopathic pulmonary fibrosis and those with pulmonary hypertension have the worst (60 to 65%). Living lobar recipients have early survival rates that are between the rates of these groups, but long-term data are not available for this population.

FUNCTIONAL OUTCOMES

Arterial blood gas levels improve markedly in unilateral and bilateral lung transplant recipients by 3 months posttransplant. In both groups, Pa_{CO_2} normalizes; in bilateral lung transplant recipients, Pa_{O_2} also normalizes. Unilateral lung recipients may continue to have mild hypoxemia but rarely require supplemental oxygen. Pulmonary function studies usually reach their maximum values for both groups between 3 and 12 months postoperatively. Unilateral graft recipients who had a preoperative diagnosis of parenchymal lung disease attain 60 to 65% of their predicted FVC and FEV_1 values. The values for bilateral lung recipients often approach normal predicted values, but these patients can have mild restrictive physiology. Diffusing capacities are usually slightly decreased in all groups. Airway hyperresponsiveness without clinically relevant asthma can be demonstrated in the majority of lung transplant recipients.

Exercise capacity has been the most interesting functional outcome observed in lung transplant recipients. With respect to nongraded exercise capacity, usually measured by 6- or 12-min walk studies, unilateral and bilateral graft recipients demonstrate marked and similar improvement in distances covered after transplantation. Typically, transplant recipients can walk 100 to 120 m/min within 6 months of transplant and are generally able to sustain this rate over time. On graded exercise studies, however, both groups achieve only 40 to 60% of predicted maximum values, with bilateral lung recipients usually performing slightly better than unilateral lung recipients. This exercise limitation has been extensively studied particularly in bilateral lung recipients. The limitation appears not to be cardiac or ventilatory but rather related to muscle deconditioning and abnormalities in skeletal muscle oxidative capacity. Rarely is the exercise limitation in these patients enough to impact on their normal daily activities or their quality of life.

POSTTRANSPLANT MANAGEMENT ISSUES

Airway Complications Technical improvements and surgical experience have greatly reduced significant anastomotic complications in lung transplant recipients. It is not uncommon to see small dehiscences of the airway in the first weeks posttransplant, but these generally heal without significant stricture. Probably fewer than 10% of patients will have stenosis severe enough to require balloon dilatation, laser resection, or a stent. When required, wire stents are most often used and are well tolerated. Late-occurring bronchomalacia, often at the anastomotic site, has also been treated with stents.

Acute Rejection A three-pronged immunosuppressive approach, which is used in most lung transplant programs, includes either cyclosporine or tacrolimus, either azathioprine or mycophenolate mofetil, and prednisone. Cytolytic induction is rarely used in these patients because of the risk of infection. Most lung transplant recipients experience at least one episode of acute rejection, usually within the first 3 months, although episodes have been reported to occur up to several years after transplantation. From 10 to 15% of patients have recurrent acute or persistent acute rejection, which predisposes them to

chronic rejection. Symptoms include a general feeling of malaise, dyspnea, and sometimes cough. Findings may include low grade fever, rales, mild hypoxemia, decreasing FVC and FEV_1values, increased white blood cell count, and ill-defined infiltrates with or without pleural effusion on chest x-ray. If a patient presents early in an episode of acute rejection, as most do, the findings are minimal and the chest radiogram is clear. Histologic diagnosis, which is the "gold standard," is routinely made by transbronchial biopsy, with a sensitivity of about 80% and a specificity approaching 100%. Bronchoscopy is also helpful in this setting to rule out infections that may have similar presentations.

Acute rejection episodes occurring early after transplantation respond in at least 80% of patients to bolus methylprednisolone. Late episodes and recurrent or persistent episodes often require both intensification of immunosuppression and changes in immunosuppressive drugs. Up to 20% of asymptomatic patients have at least one episode of acute rejection detected by surveillance transbronchial biopsy in the first 2 years posttransplant. It is not clear whether asymptomatic rejection requires therapy, as the impact on outcome is unknown. Thus, the use of surveillance bronchoscopy and the treatment of asymptomatic rejection vary considerably from institution to institution, and there are at present no clear guidelines in this area.

Bronchiolitis Obliterans Bronchiolitis obliterans is both the primary manifestation of chronic rejection and the most feared complication in lung transplant recipients. It occurs to some degree in at least 50% of survivors by 5 years posttransplant and is a factor in more than one-third of late deaths. Although it can occur as early as 2 months posttransplant, the onset is more often at least 6 months and the mean onset is from 1 to 2 years after surgery. The precipitating factors and initiating events in bronchiolitis obliterans are topics of intensive research both in transplant recipients and in several animal models. Those factors most consistently associated with the process include the numbers and severity of acute rejection episodes and episodes of cytomegalovirus (CMV) pneumonia, but not clearly CMV infection alone. Other factors with weaker associations include HLA mismatches, other viral infections, and the development of anti-HLA antibodies. Clinically, the onset of this process is often subacute, with a very gradual onset of dyspnea and fatigue or malaise, often accompanied by viral-type symptoms or dry cough. It can also be asymptomatic and detected by routine pulmonary function studies that show, initially, a decrease in the FEF_{25-75}, often followed one to several months later by decreasing FEV_1. This insidious development of small airway obstruction is often well established before it is clinically recognized, and for that reason frequent pulmonary function testing is recommended for lung transplant recipients. Chest radiograms are usually normal, but even early in the disease expiratory computed tomography (CT) scans show a mottled appearance with peripheral hyperlucency. Transbronchial biopsy is very specific but not sensitive in diagnosis, but patients usually undergo at least one bronchoscopy at the onset of disease to attempt histologic documentation and to rule out possible infections. Because of the difficulty in histologic diagnosis, a typical clinical picture in the absence of other etiology is considered sufficient to establish a diagnosis of *bronchiolitis obliterans syndrome*. The progression of this complication can be very rapid, with early death; but more often it is one of a gradually decreasing FEV_1over months to years, which in the later stages is frequently accompanied by bronchomalacia, proximal bronchiectasis, and recurrent pseudomonal or other infections.

Effective treatment remains evasive. A few immunosuppressive protocols tried in small numbers of patients have been found to "stabilize" pulmonary function, but improved function is unusual. Likely, by the time the process is recognized in most patients, fibrotic obliteration of the airway is already present; the key to treatment may lie in identifying markers of incipient disease and much earlier intervention.

Infections Infections rank second only to rejection as a cause of morbidity in lung transplant recipients and are the most common cause of mortality, accounting for one-third of all deaths in both the early and the late posttransplant periods. The transplanted lung may be uniquely vulnerable to infection because of impaired mucociliary clearance, loss of cough reflex, and other poorly defined local factors. In addition, the donor lungs are often colonized with organisms that are transmitted directly to the immunosuppressed recipient. Early series reported that at least 60% of lung recipients early in their course develop infections requiring treatment. Now the extensive use of broad antibacterial, antifungal, anti-*pneumocystis*, and antiviral prophylaxis, often maintained for at least 3 months postoperatively, seems to have reduced the early infective morbidity and mortality.

Infections with paramyxoviral organisms, adenovirus, and influenza A have now been well documented and have an overall death rate of about 20%. The role of antiviral therapy is unclear. The most lethal infections are those with invasive fungal organisms, particularly those caused by *Aspergillus* species, which have been reported to colonize in 20 to 50% of recipients. Invasive disease caused by these organisms can vary from ulcerative bronchitis to localized parenchymal infiltrates to empyema to disseminated disease. *Aspergillus* is particularly likely to be problematic when patients require increased immunosuppression or have other complications; one study has reported an increased rate of invasive disease in the native lung of unilateral lung transplant recipients.

The highest risk periods for infection are in the first few months posttransplant and late after transplant if bronchiolitis obliterans or other vital organ dysfunction, e.g., renal failure, develops. Since it may be very difficult to distinguish infection from rejection in the early posttransplant period, bronchoscopy with appropriate biopsies and cultures is often necessary to establish a diagnosis.

Immunosuppressive and Medical Complications Medical complications related to immunosuppression, to the underlying diagnosis, or to aging are major causes of morbidity in long-term survivors of lung transplantation and account for up to 10% of late deaths. Current immunosuppressive regimens with cyclosporine or tacrolimus cause some nephrotoxicity in virtually all patients. Although few progress to renal failure, hypertension and hyperlipidemia are common. Neurotoxicity, including delirium, headaches, seizures, and, occasionally, strokes, has been reported in up to 20% of patients. Osteoporosis occurs in more than half the patients, and vertebral compression fractures are common. Other problems include thromboembolic disease, gastric complications (especially gastroparesis), hyperglycemia, and increased rates of malignancy.

Posttransplant lymphoproliferative disorders associated with Epstein-Barr virus occur in

5 to 10% of lung transplant recipients. Nearly all occur within the first year after transplant. Recent data suggest a much higher incidence of this disease in patients who are Epstein-Barr naive and who receive an Epstein-Barr positive graft. Treatment for this disorder, reported to have an approximate 50% survival rate, is usually reduced immunosuppression and antiviral and anti-B lymphocyte drugs. Survivors often develop bronchiolitis obliterans.

Recurrence of Underlying Disease Several different underlying diseases may recur in lung transplant recipients. These diseases include sarcoidosis, lymphangioleiomyomatosis, giant cell interstitial pneumonia, panbronchiolitis, eosinophilic granuloma, bronchoalveolar cell carcinoma, and desquamative interstitial pneumonia.

(Bibliography omitted in Palm version)

Back to Table of Contents

PART TEN -DISORDERS OF THE KIDNEY AND URINARY TRACT

268. DISTURBANCES OF RENAL FUNCTION - *Robert M. Brenner*, *Barry M. Brenner*

Near constancy of the composition of the internal environment, including the volume, tonicity, and compartmental distribution of the body fluids, is essential to survival. With normal day-to-day variations in the intake of food and water, preservation of the internal environment requires the excretion of these substances in amounts that balance the quantities ingested. Although losses from intestines, lungs, and skin contribute to this excretory capacity, the greatest responsibility for solute and water excretion is borne by the kidneys.

The kidneys regulate the composition and volume of the plasma water. This, in turn, determines the composition and volume of the entire *extracellular* fluid compartment. Through the continuous exchange of water and solutes across all cell membranes, the kidneys influence the *intracellular* fluid compartment as well. These functions are served by a variety of physiologic mechanisms that enable individuals to excrete excesses of water and nonmetabolized solutes contained in the diet, as well as the nonvolatile end products of nitrogen metabolism, such as urea and creatinine. Conversely, when faced with deficits of water or solute, excretion of water or specific solute(s) is curtailed via appropriate mechanisms for renal conservation, reducing the likelihood of volume or solute depletion. The purpose of this**chapter**is to review the excretory functions of the kidney and to examine how these functions are affected by chronic renal disease.

EFFECTS OF NEPHRON LOSS ON RENAL EXCRETORY MECHANISMS

The volume of urine excreted (averaging 1.5 L/d or roughly 1 mL/min) represents the sum of two large, directionally opposite processes -- namely, *ultrafiltration* of 180 L/d or more of plasma water (or 125 mL/min) and *reabsorption* of more than 99% of this filtrate by transport processes in the renal tubules. While renal blood flow accounts for about 20% of resting cardiac output, the kidneys comprise only about 1% of total body weight. This disproportionate allocation of cardiac output, greatly exceeding blood flow per gram of brain, heart or liver, is required for the process of ultrafiltration.

GLOMERULAR ULTRAFILTRATION

Urine production begins at the glomerulus where an ultrafiltrate of plasma is formed. The rate of glomerular ultrafiltration (glomerular filtration rate, GFR) is governed chiefly by forces favoring filtration on the one hand (hydraulic pressure in the glomerular capillaries) and forces opposing filtration on the other (the sum of hydraulic pressure in Bowman's space and colloid osmotic pressure in the glomerular capillaries). The rate of glomerular plasma flow and the total surface area of the glomerular capillaries are also determinants of GFR. Decreased GFR can therefore be expected when (1) glomerular hydraulic pressure is reduced (as in circulatory shock); (2) tubule (hence Bowman's space) hydraulic pressure is elevated, as in urinary tract obstruction; (3) plasma colloid osmotic pressure rises to high levels (hemoconcentration due to severe volume depletion, myeloma, or other dysproteinemias); (4) renal, and hence glomerular, blood flow is reduced (severe hypovolemia, cardiac failure); (5) permeability is reduced

(diffuse glomerular disease); or (6) filtration surface area is diminished, through focal or diffuse nephron loss in progressive renal failure.

The glomerular capillary wall is specially adapted to allow passage of extremely large volumes of water while retaining all but the smallest solute molecules. Molecules the size of inulin (approximately 5200 mol wt) pass freely across the glomerular filtration barrier, appearing at approximately the same concentration in Bowman's space as in plasma. The passage of solutes across the glomerular barrier decreases progressively with increasing molecular size such that, as the molecular weight of albumin is approached, most of the solute is retained in the plasma. Albumin, a polyanionic molecule in plasma, is further retarded at the glomerular filtration barrier by *electrostatic forces* imparted by negatively charged cell-surface molecules on the epithelial foot processes that form the *filtration slits* and the *slit diaphragms*. With disruption of these structural and electrostatic barriers, as in many forms of glomerular injury (Chaps. 273 to 275), large quantities of plasma proteins gain access to the glomerular filtrate.

Glomerular Adaptations to Nephron Loss With loss of nephron mass, the remaining functional (or least injured) nephrons tend to hypertrophy and take on an increased workload so that the overall loss of function is minimized. For example, a patient with a unilateral nephrectomy loses one-half of the nephron mass, resulting in a 50% reduction inGFR at the time of surgery. However, the GFR in the remaining kidney begins to increase after 1 or 2 weeks, and within several months GFR may rise to 80% of the preoperative value. This indicates that the GFR of the individual remaining nephrons has increased above normal, a state known as *hyperfiltration*. Increases in single-nephron GFR may be achieved by renal hemodynamic adjustments (increased glomerular plasma flow and increased glomerular capillary hydraulic pressure), which augment the forces driving ultrafiltration, and by glomerular hypertrophy, which increases the maximum surface area available for filtration. These structural adaptations are evident from the enlargement of glomeruli (and tubules) seen on histologic sections from people with single kidneys. Similar structural changes are observed in kidneys damaged by chronic disease processes; foci of hypertrophied glomeruli and tubules are interspersed with areas of atrophic or scarred parenchyma. Although direct measurements of single-nephron GFR cannot be made in humans, it is reasonable to conclude that focal nephron enlargement as occurs in chronically diseased kidneys generally signifies focally increased single-nephron GFR, and that these dynamic adaptations represent compensatory adjustments for the effects of nephron loss through disease.

Glomerulotubular Balance The close integration of glomerular and tubular functions (*glomerulotubular balance*) seen in chronic renal failure (CRF) supports the notion that progressive nephron obliteration is the usual mode ofGFRreduction in CRF. Preservation of glomerulotubular balance until the terminal stages of CRF is fundamental to the *intact-nephron hypothesis*, which states that as CRF advances, kidney function is supported by a diminishing pool of functioning (or hyperfunctioning) nephrons, rather than relatively constant numbers of nephrons, each with diminishing function. This concept has important implications for the mechanisms of disease progression in CRF. A considerable amount of evidence suggests that nephrons subjected to increased excretory burdens for prolonged periods actually sustain injury as a result of these adaptations: thus the cost of these compensatory adaptations to

nephron loss may ultimately be relentless destruction of the remaining nephron pool.

The magnitude of the single-nephron hyperfiltration induced by loss of 50% of the total nephron mass usually has no serious adverse clinical consequences, even when sustained over two to three decades. When more than 50% of the total nephron mass is lost, however, as in renal-sparing surgery for bilateral trauma or neoplasm or from a renal disease whose activity has abated, the remaining nephrons are forced to the limits of their compensatory capacity. While these adaptations achieve remarkable short-term success at offsetting the tendency forGFR to fall, over time, proteinuria and focal and segmental glomerulosclerosis develop, the more so where greater amounts of nephrons are lost or removed. As a result, a progressive decline in GFR ensues. Experimental study of the processes that advance glomerular injury show that the adverse long-term consequences of severe nephron deficits are invariably preceded by increases in glomerular capillary hydraulic pressure (glomerular capillary hypertension), glomerular hyperperfusion, and hypertrophy. Interventions directed against these compensatory and maladaptive responses can greatly ameliorate the subsequent development of renal failure. In particular, drugs (e.g., angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers) and other interventions (such as dietary protein restriction) that lower glomerular pressure can slow the rate of progression of experimental and human renal disease. In the absence of such interventions, more and more glomeruli cease to function through advancing glomerulosclerosis and disruption of tubule structure and function, leading eventually to total loss of GFR (i.e., end-stage renal disease). This *final common pathway* for chronic renal injury helps to explain the observed progressive nature of chronic renal failure resulting from many different kidney diseases.

Biologic Consequences of Sustained Reductions in GFR Although nephron loss can proceed, to some extent, without equivalent loss ofGFR due to the compensatory mechanisms described above, determination of the total GFR of both kidneys remains the most reliable clinical index of overall excretory function. The effects of impaired GFR are to reduce the total rate of delivery of solute into the glomerular filtrate. When accompanied by comparably reduced rates of urinary excretion, *retention* and *accumulation* of the unexcreted solute occurs, resulting in increased concentrations of the substance in the plasma and other body fluids.

Figure 268-1 depicts the major types of response to impairedGFR. The degree of reduction in total GFR is plotted on the abscissa, expressed as a percentage of normal (100%). The renal handling of most solutes normally present in glomerular filtrate conforms to one of three patterns. Curve *A* describes the pattern with substances such as creatinine and urea that normally depend largely on glomerular filtration for urinary excretion; i.e., secretion contributes little to overall excretion. Therefore, as illustrated, gradual reductions in GFR are accompanied by progressive increases in plasma levels of creatinine, urea, and other substances normally excreted primarily by filtration.

The clinical course ofCRFusually also approximates the pattern described by curve *A*. Patients with CRF usually pass from a long asymptomatic period of "compensation" to a more accelerated and clinically overt terminal phase. In other words, despite chronic injury leading to destruction of more than 50% of nephrons, plasma elevations of creatinine and urea may still lie within the normal limits for these substances. With

further nephron loss and reduction inGFR, however, the limits of renal reserve are exceeded and continued accumulations of curve *A*-type solutes lead to abnormally elevated plasma concentrations (Fig. 268-1). Because some of these retained solutes are thought to exert "toxic" effects on all organ systems, clinical manifestations of CRF may now become apparent. Consequently, in patients with substantial reductions in nephron mass but near-normal plasma creatinine, overt uremia may be precipitated by a modest additional decline in GFR.

The accumulation of curve *A*-type solutes with chronic loss of renal function proceeds until external balance is restored, i.e., intake and/or production rates exactly match excretion rates. In the case of creatinine, for example, assuming a constant rate of creatinine production, a 50% reduction inGFRresults in an approximate doubling of the plasma creatinine concentration. The latter restores the filtered load of creatinine (i.e., the product of GFR and plasma creatinine concentration) to normal, and the urinary excretion rate once again is equivalent to creatinine production. Since creatinine secretion contributes only slightly, elimination of the retained creatinine is not possible and the plasma concentration remains twice normal. With further loss of GFR, elevations in plasma creatinine are compounded by loss of nephron excretory function and creatinine retained as the result of earlier nephron destruction (Fig. 268-1). *In practice, so long as the net rates of acquisition and production (i.e., liver function and muscle mass) remain reasonably constant, the inverse relationship between plasma concentrations of solutes such as creatinine and urea and GFR is sufficiently reliable to serve as clinical indices of GFR.* However, where muscle mass is low, as with severe weight loss, unremarkable plasma levels of creatinine may belie substantial reductions in GFR.

In contrast to solutes of the curve *A* type, plasma levels of phosphate (PO_4^{3-}), urate, and potassium (K^+) and hydrogen (H^+) ions usually do not rise until theGFRfalls to a small percentage of normal. With progressive renal failure this pattern of response (curve *B* inFig. 268-1) reflects the participation of tubule transport mechanisms in the excretion of these substances. In other words, *as GFR declines, the tubules facilitate greater elimination of these substances, by enhancing secretion and/or by diminishing reabsorption, so that a greater fraction of the filtered load is excreted.* Plasma levels of curve *B*-type solutes, therefore, rise less than those of curve *A* because, with progressive reductions in GFR, *excretion rate per nephron* and therefore *fractional excretion* both increase. Eventually, however, with further loss of GFR, enhanced fractional excretion can no longer mitigate the reduction in net filtered load of these solutes and plasma levels rise (Fig. 268-1). For urate, PO_4^{3-}, and K^+, at least, increased fractional excretion serves to maintain normal plasma levels until GFR falls to less than one-fourth of normal.

Finally, for certain solutes, such as sodium chloride (NaCl), plasma concentrations remain normal throughout the course ofCRF, despite unrestricted intake of these substances (curve *C* inFig. 268-1). The compensatory mechanism required to achieve this represents a fundamental adaptation to chronic renal injury. To illustrate the magnitude of this adaptation, it is useful to compare the excretion of sodium (Na^+) in a normal individual (GFR of 125 mL/min) with that of a patient with advanced renal failure (GFR of 2 mL/min). Both individuals consume a conventional diet containing 7 g/d of salt (120 mmol Na^+). With a normal serum Na^+concentration of 140 mmol/L, external

Na^+balance is achieved by excreting approximately 0.5% of the filtered load. By contrast, for external balance to be maintained in the patient with CRF, fractional excretion of Na^+ must rise to 30%. In other words, *to maintain external* Na^+balance, the same amount of Na^+ must be excreted into the urine each day in the patient with CRF as in the normal individual. Given the drastic reduction in GFR in CRF, external balance can only be maintained by marked adaptations in the reabsorptive processes in surviving tubules. In this manner, a progressively larger fraction of the filtered load escapes reabsorption and appears in the final urine. In short, *the rate of excretion of* Na^+per surviving nephron increases in inverse proportion to the composite GFR in surviving nephrons.

ADAPTATIONS IN TUBULE TRANSPORT MECHANISMS IN RESPONSE TO NEPHRON LOSS

Despite progressive nephron loss, many mechanisms that regulate renal solute and water balance differ only quantitatively, and not qualitatively, from those that operate normally. Thus, glomerulotubular balance is maintained. The most important of these mechanisms are considered below.

TUBULAR TRANSPORT OF SODIUM CHLORIDE AND WATER

Most of the filtered water and sodium salts are reabsorbed by the tubules, leaving small and variable amounts, equivalent on average to the quantities ingested, to reach the final urine. About two-thirds of the glomerular ultrafiltrate is reabsorbed in the *proximal tubule* with little change in the osmolality or Na^+concentration of the unreabsorbed fraction (Fig. 268-2). In other words, fluid reabsorption in the proximal tubule is nearly *isosmotic* and is coupled to the active transport of Na^+. Since chloride (Cl^-) and bicarbonate (HCO_3^-) are the primary anions in the extracellular fluid, they constitute the main solutes that accompany Na^+reabsorption in the renal tubules. In the earliest portion of the proximal tubule, bicarbonate is the principal anion that accompanies the reabsorption of Na^+. This process occurs via a Na^+/H^+exchanger at the luminal brush border and is dependent on the activity of carbonic anhydrase. Glucose, amino acids, and other organic solutes (e.g., lactate) are also extensively reabsorbed in the proximal tubule by cotransport mechanisms that link the cellular entry of these organic molecules with Na^+. The coupling of water absorption (i.e., volume) with solute absorption appears to be dependent upon three processes. First, given the remarkably high water permeability of this segment, very small transepithelial osmolality differences, i.e., *luminal hypotonicity* of the order of 2 to 3 mosmol/L produced by solute absorption, could drive water absorption. Second, due to preferential absorption of $HCO_3^-$and organic solutes in the early portions of the proximal tubule, the concentrations of these substances decrease along the proximal tubule while that of chloride increases. Volume reabsorption would then occur if the diffusion of Na^+ and Cl^- down their respective electrochemical gradients across the proximal tubule epithelium occurred more easily than the back-diffusion of sodium bicarbonate into the lumen, creating an *effective osmotic pressure gradient*. Finally, *lateral interstitial space hypertonicity* produced by differences in the rates at which solutes are transported into the spaces or exit them by diffusion may also contribute to the coupling of water and solute reabsorption.

Reabsorption of Fluid from Proximal Convoluted Tubules This is sensitive to

Starling forces, i.e., the hydraulic and colloid osmotic (or oncotic) pressures acting across the walls of the peritubular capillaries. Because the plasma proteins in glomerular capillaries are concentrated by ultrafiltration, oncotic pressure rises along the glomerular capillary network. This step-up in oncotic pressure is transmitted largely unchanged to the first branches of the peritubular capillaries via the efferent arterioles. These resistance vessels cause a substantial drop in hydraulic pressure, however, so that when the plasma reaches the peritubular capillaries, oncotic pressure greatly exceeds hydraulic pressure. The Starling forces are therefore oriented in an *uptake* mode, in contrast to their configuration at the glomerulus where hydraulic pressure exceeds oncotic pressure, favoring *filtration.* The extent to which oncotic pressure exceeds hydraulic pressure in the peritubular capillary network modulates the overall rate of fluid absorption by the peritubular capillaries. Therefore, when peritubular capillary oncotic pressure falls, or hydraulic pressure rises, uptake of fluid by these capillaries is reduced. As a result, fluid is retained in the interstitial space, tending to increase hydraulic pressure, ultimately retarding the egress of fluid from the lateral intercellular channels. Without an adequate route of drainage, fluid in the intercellular channels leaks back into the tubule lumen, thereby *diminishing net fluid reabsorption* from this tubule segment. The opposite occurs in states where peritubular oncotic pressure is increased (increased filtration fraction) or hydraulic pressure is decreased (enhanced efferent arteriolar tone). Under these circumstances, peritubular capillary uptake of reabsorbate is augmented, leading ultimately to *enhanced net fluid reabsorption* by the proximal tubule. Although physical factors appear to be the major determinants of fluid reabsorption in the proximal tubule, hormones (e.g., angiotensin II) may also modulate fluid reabsorption directly, by enhancing luminal Na^+ entry into proximal tubule cells via an apical Na^+/H^+ exchanger.

The Limbs of Henle's Loop In contrast to the proximal tubule, active outward transport of Na has not been established for the *thin ascending limb of Henle's loop*. However, passive outward salt transport does occur, as indicated in Fig. 268-2. In the next nephron segment, the *medullary thick ascending limb of Henle*, the concentration of NaCl is reduced as fluid traverses this segment. Here Cl^- absorption occurs by an active process involving a $Na^+:K^+:2Cl^-$ cotransport mechanism in the luminal membrane, with one-half of Na^+ absorption proceeding passively, driven by the lumen positive transepithelial voltage difference. This cotransporter is the site of action of the powerful loop diuretics and mutations give rise to Bartter's syndrome. Since the ascending limb of Henle is impermeable to water, net NaCl reabsorption generates a hypotonic tubule fluid and gives rise to the high NaCl concentration of the outer medullary interstitium (Fig. 268-2). In certain animals, arginine vasopressin (AVP; also called ADH) enhances NaCl absorption in the medullary portion of the thick ascending limb, but whether this occurs in humans is uncertain.

Distal Tubule The fluid leaving the thick ascending limb of Henle is normally of low NaCl concentration, a characteristic independent of the organism's hydration or dietary status. In the *distal tubule*, water reabsorption is variable, depending on the state of hydration or, specifically, on the presence or absence of AVP in plasma. In the absence of AVP, this and more distal nephron segments are impermeable to water, so that hypotonic fluid entering this segment is excreted as *dilute urine.* Indeed, continued salt reabsorption along the distal convoluted tubule (DCT) and connecting tubule segments, a process that can be inhibited by the thiazide classes of diuretics, results in further

dilution of the urine. In the presence of AVP, the permeability of these nephron segments to water increases. This is made possible by the insertion of proteins known as *aquaporins* into the luminal cell membrane of DCT cells. These proteins facilitate water movement from the low osmolality environment of the DCT lumen into the higher osmolality of the medullary interstitium, thereby contributing to the creation of a concentrated final urine. NaCl continues to be reabsorbed from the tubule lumen against moderately steep chemical and electrical gradients. The reabsorption of NaCl at the collecting tubule is enhanced by *aldosterone.*

Collecting Tubules and Ducts The *cortical collecting tubule* possesses a low permeability to water in the absence ofAVP, whereas permeability increases in the presence of this hormone. The sensitivity of this segment to AVP appears to be more pronounced than that of the DCT. As with the DCT, the cortical collecting tubule is capable of active reabsorption of NaCl and its stimulation by aldosterone.

The terminal segment of the distal nephron is the highly branched *papillary collecting duct*. Continued electrolyte transport in this segment results in the large ion concentration differences that normally exist between urine and plasma. As in the cortical collecting tubule, Na^+transport appears to be active, since reabsorption proceeds against sizeable electrochemical gradients. The rate of Na^+transport in this segment depends on the load of Na^+delivered from more proximal segments and is also affected by aldosterone. The permeability to water is also increased markedly in the presence ofAVP.

Effects of Nephron Loss on Sodium Chloride Transport in Surviving Nephrons With progressive nephron loss, *maintenance of external balance for NaCl requires that fractional salt excretion increases in concert with the decline inGFR*. Several mechanisms contribute to this adaptive increase in fractional Na^+excretion. With loss of functioning nephron units, peritubular capillary Starling forces are presumably altered in directions that serve to reduce proximal tubule reabsorption of NaCl and water. For example, a rise in peritubular capillary hydraulic pressure, which tends to inhibit net proximal fluid reabsorption, might be anticipated with systemic hypertension, a common feature of chronic renal failure. Similarly, reductions in peritubular capillary oncotic pressures may be anticipated due to reductions in both filtration fraction and hypoalbuminemia.

Aldosterone, which normally exerts a potent influence on tubule transport, probably does not figure prominently in reducing fractional Na^+excretion, since aldosterone levels are seldom reduced inCRF. Furthermore, external Na^+balance is preserved in bilaterally adrenalectomized dogs on fixed replacement doses of mineralocorticoid. Yet another factor contributing to the suppression of fractional NaCl reabsorption in CRF may relate to the retention of various organic solutes asGFRdeclines.

Several factors that regulate NaCl transport across tubules under resting conditions are also likely to contribute to the enhanced fractional excretion of salt in renal insufficiency. Atrial natriuretic peptides are released from the heart in response to elevated cardiac (atrial) filling pressures as seen with increased plasma volume or atrial tachyarrhythmias. These peptides affect natriuresis by reducing net Na^+reabsorption through complementary actions on Na^+transport in the collecting duct and by altering

Starling forces in the adjacent vasa recta. The vascular actions of natriuretic peptides may also extend to glomerular hemodynamics, with afferent arteriolar vasodilatation contributing to increased single-nephronGFR and hence an increase in the amount of Na^+filtered. Other modulators of tubule transport processes may also contribute to increased single-nephron natriuresis in the setting of reduced renal mass or nephron loss. Vasodilator prostaglandins are present at increased plasma levels in CRF, as are other inhibitors of transport, including inhibitor(s) of the Na^+,K^+-ATPase. This latter factor has not yet been fully characterized; whether its presence represents a homeostatic adaptation for maintenance of fluid balance or an unregulated accumulation of a toxin remains uncertain.

Serum and urine from patients with uremia contain factors capable of experimentally inhibiting NaCl transport across frog skin, toad bladder, and rat renal tubule. Accumulation of natriuretic factors in uremia may not be without cost; the "trade-off" for maintenance of external Na^+balance is the possibility of generalized abnormalities occurring in Na^+transport across cell membranes, which often occur in advanced renal failure (Chap. 270).

The obligatory high rate of solute excretion per surviving nephron (so-called osmotic diuresis due to urea and other retained solutes) also contributes to enhancing fractional NaCl excretion, much as occurs in normal individuals after the administration of mannitol or other nonreabsorbable solutes. Finally, certain forms ofCRF are associated with unusually large losses of salt in the urine. These *salt-wasting nephropathies* include chronic pyelonephritis and other tubulointerstitial diseases (Chap. 277) as well as polycystic and medullary cystic diseases. These disorders have in common greater destruction of medullary and tubulointerstitial, rather than cortical and glomerular, portions of the renal parenchyma. Preferential impairment of tubule reabsorptive function, rather than a primary reduction in glomerular filtration, may, therefore, underlie the salt-losing tendency in these disorders. Clinical derangements that alter renal handling of NaCl in CRF (including hypo- and hypervolemia, hypertension, etc.) are considered in Chap. 270.

EFFECTS OF NEPHRON LOSS ON WATER REABSORPTION IN SURVIVING NEPHRONS

As with NaCl, there is a progressive increase in the fractional excretion of water with advancing renal insufficiency, so that external water balance can be maintained even with a totalGFR of 5 mL/min or less. The adaptations of water handling by the diseased kidney are of importance in the defects in urinary concentration and dilution and hence the polyuria, nocturia, and tendency to develop water overload encountered inCRF(Chap. 47). To appreciate the mechanisms involved, the responses of a normal and a uremic individual maintaining external water balance need to be considered. Assuming both individuals have the same dietary and fluid intakes, total solute and volume excretion in both should be identical as well. If the *obligatory solute load* to be excreted by each is 600 mmol/d (600 mosmol/d) and the urine osmolality is 300 mmol/kg water (300 mosmol/kg), a urine volume of 2 L/d will be required to excrete the total solute. If the GFR in normal and uremic individuals totals 180 and 4 L/d, respectively, urinary volume excretion of 2 L/d represents excretion of slightly more than 1% of the total glomerular filtrate in the normal subject compared with 50% in the uremic

patient. Since the range of urine osmolalities that the diseased kidney can achieve [250 to 350 mmol/kg (250 to 350 mosmol/kg)] is narrower than in the normal kidney [40 to 1200 mmol/kg (40 to 1200 mosmol/kg)], the individual with normal function is able to excrete the obligatory daily solute load of 600 mmol (600 mosmol) in as little as 500 mL urine per day or as much as 15 L/d, compared with the narrower range in renal insufficiency, from about 1.7 to 2.4 L/d.

InCRF, the limited capacity to concentrate the urine often correlates with other measures of impaired renal function. Isosthenuria (urine of similar osmolality to plasma) is therefore an almost universal finding when theGFRfalls below 25 mL/min. At this level of GFR and below, urine osmolality does not rise even when supraphysiologic doses ofAVP are administered, suggesting that the concentrating defect relates to impaired concentrating capacity in surviving nephrons. The associated increased fractional excretion per nephron of a variety of solutes produces an obligatory water loss (solute diuresis) at roughly isotonic proportions. Consequently, formation of a concentrated urine is prevented. Disease-induced abnormalities of the architecture of the renal medulla (loops of Henle, vasa recta), aberrations in medullary blood flow, and defective transport of NaCl in the ascending limb of Henle also contribute to this defect in urine concentration.

Since patients withCRF are unable to excrete concentrated or dilute urine, they must have access to adequate, and to some extent, relatively constant amounts of water per day to ensure that they have adequate water to eliminate total daily solute loads. For this reason, restriction of fluid intake may be hazardous in patients with CRF. Likewise, impairment of diluting capacity may prevent many patients from excreting excess ingested fluid. The consequences of the abnormal patterns of water excretion, and the attendant susceptibilities to develop hypo- and hypernatremia, are considered inChaps. 49 and270.

TUBULE TRANSPORT OF PHOSPHATE WITH NORMAL AND REDUCED NEPHRON MASS

Under normal physiologic conditions, about 80 to 90% of phosphate is reabsorbed, mainly in the proximal tubule. *Parathyroid hormone* (PTH), by augmenting phosphate excretion via inhibition of this proximal reabsorptive process (Chap. 340), plays a central role in phosphate homeostasis. When dietary phosphate intake increases, a *transient* rise in plasma phosphate concentration is usually observed. This results in a similarly transient reduction in the plasma ionized calcium level (due largely to deposition of calcium phosphate in bone), which is sensed by a specific receptor on parathyroid cells, stimulating PTH secretion. By enhancing fractional phosphate excretion, PTH restores external phosphate balance and normophosphatemia. This enables plasma ionized calcium levels to return to normal, thereby removing the stimulus to PTH release and restoring the phosphate control system to the original steady state.

With advancing renal failure and constant dietary intake of phosphate, external phosphate balance is achieved by progressive reduction in fractional phosphate reabsorption. EnhancedPTHsecretion is an important determinant of this phosphaturic response. With succeeding decrements in totalGFR, the amount of phosphate filtered by surviving glomeruli is reduced, leading to transient phosphate retention and,

therefore, a rise (albeit small) in plasma phosphate concentration. This leads to a small, reciprocal decline in plasma levels of ionized calcium and a corresponding increase in PTH secretion. Although the phosphaturic response of surviving tubules to this elevation in circulating PTH restores plasma phosphate and calcium to normal levels (at least in the "compensated" stage ofCRFdescribed by curve *B* in Fig. 268-1), the new steady-state conditions are only achieved at the cost of *persistently elevated plasma PTH levels*. With progressive reductions in GFR, the process is repeated, resulting in substantially elevated PTH levels.

Alterations in Vitamin D Metabolism These alterations also contribute to elevatedPTHlevels inCRF. The kidney is normally the major site of *conversion of vitamin D to its active metabolites*. As discussed in Chap. 340, vitamin D, synthesized in skin or acquired in the diet, undergoes initial hydroxylation in the liver to form 25-hydroxyvitamin D [25(OH)D]. The kidney is the site of a second important conversion to 1,25-hydroxyvitamin D [1,25(OH_2)D]. This active form of vitamin D acts directly on the parathyroid gland to suppress PTH secretion as well as to enhance intestinal absorption of calcium and phosphate resorption and promote resorption of these ions from bone. In addition, 1,25(OH)D_2probably opposes the phosphaturic actions of PTH in the renal tubule by augmenting, rather than diminishing, phosphate reabsorption. With advancing renal disease, nephron loss reduces the renal capacity for vitamin D hydroxylation; phosphate retention also impairs this reaction. Not only are the circulating levels of 1,25(OH)D_2diminished in uremia, but the receptors that mediate its action at the parathyroid gland are also diminished. These two effects remove inhibitory influences on PTH secretion, leading again to increased plasma PTH levels. Reduction in circulating 1,25(OH)D_2levels, by suppressing intestinal calcium absorption, contributes to the development of the hypocalcemia and hyperparathyroidism of CRF (Chap. 270).

Hyperparathyroidism in Chronic Renal Failure At least two additional processes are thought to contribute to hyperparathyroidism inCRF. One relates to resistance of bone to the calcemic effect ofPTH in uremia. This resistance necessitates a higher level of PTH to demineralize bone and maintain the plasma calcium concentration. The other derives from the finding that reductions in renal mass impair the kidneys' capacity to degrade circulating PTH. Ultimately, however, phosphate conforms more to a curve *B*-rather than a curve *C*-type pattern inFig. 268-1, and phosphate retention occurs when theGFRfalls below about 25 mL/min, signifying that these latter forms of adaptation play limited roles.

SincePTHexerts major effects on bone as well as renal tubules, the external balance of phosphate inCRF is achieved at the expense of elevated PTH levels, which, in turn, account for many of the bone changes of renal osteodystrophy (i.e., *secondary hyperparathyroidism*; Fig. 270-1). In support of this *trade-off hypothesis*, when dietary phosphate intake is reduced in proportion to the reduction inGFR in animals with CRF, external balance of phosphate no longer requires augmentation of fractional phosphate excretion in surviving nephrons. Accordingly, circulating levels of PTH no longer rise, and the bone changes of secondary hyperparathyroidism are diminished, if not prevented.

HYDROGEN AND BICARBONATE TRANSPORT WITH NORMAL AND REDUCED RENAL MASS

As discussed in Chap. 50, the pH of extracellular fluid is normally maintained within a narrow range (7.36 to 7.44) despite day-to-day fluctuations in the quantity of acids added to the extracellular fluid from dietary and metabolic sources (approximately 1 mmol H^+ per kilogram of body weight per day). These acids consume buffers from both extracellular and intracellular fluid, of which HCO_3^- is the most important in the intracellular compartment. Such buffering minimizes changes in pH. Long-term effectiveness of the HCO_3^- buffer system, however, requires mechanisms for replenishment, otherwise unrelenting acquisition of nonvolatile acids from dietary and metabolic sources would ultimately exhaust buffering capacity, culminating in fatal acidosis. The kidneys normally function to prevent this eventuality by *regenerating* bicarbonate, thereby maintaining plasma concentrations of HCO_3^-. In addition, the kidneys also *reclaim* HCO_3^- in the glomerular ultrafiltrate. Reclamation of filtered HCO_3^- takes place largely in the proximal tubule and, under normal circumstances, is virtually complete below a critical plasma HCO_3^- concentration -- the threshold concentration -- which in humans is normally about 26 mmol/L, identical to the concentration of HCO_3^- in plasma. As a consequence, HCO_3^- wastage is prevented. Alternatively, when plasma HCO_3^- rises above this threshold, reabsorption becomes less complete, allowing escape of excess HCO_3^- into the final urine, which restores the plasma HCO_3^- towards normal levels. Despite complete reabsorption of HCO_3^-, metabolic acidosis would still ensue if HCO_3^- consumed in buffering nonvolatile acids were not constantly regenerated.

The *reabsorption* of filtered HCO_3^- occurs by the following mechanism. Filtered bicarbonate combines with H^+ secreted from proximal tubule cells via the Na^+/H^+ exchange, to form carbonic acid (H_2CO_3). Dehydration of carbonic acid under the influence of *luminal* carbonic anhydrase yields H_2O and CO_2, which is free to diffuse from lumen to peritubular blood. In the proximal tubule cell, the OH^- left behind by the H^+ secretion reacts with CO_2, under the influence of *intracellular* carbonic anhydrase, forming HCO_3^-. This ion is transported across the contraluminal proximal tubule cell membrane, via an electrogenic Na/HCO_3^- cotransporter, to reenter the extracellular HCO_3^- pool. The net result is *reclamation of a filtered bicarbonate ion*. Secreted H^+ is also free to react with nonbicarbonate buffers [e.g., phosphate or ammonia (NH_4^+)] in the tubule lumen, and hydrogen ions are excreted in these forms in the final urine. Again, the OH^- left behind in the proximal tubule cell from H^+ secretion reacts with CO_2, forming bicarbonate -- also representing *regeneration of an HCO_3^- ion*.

Hydrogen ions in the urine are bound to filtered buffers (e.g., phosphate) in amounts equivalent to the amounts of alkali required to titrate the pH of the urine up to the pH of the blood (the so-called titratable acid). It is not usually possible to excrete all the daily acid load in the form of titratable acid due to limits of urinary pH. Metabolism of glutamine by proximal tubule cells to yield ammonium (ammoniagenesis) serves as an additional mechanism for H^+ elimination and bicarbonate regeneration. Glutamine metabolism forms not only NH_4^+ (i.e., NH_3 plus H^+) but also HCO_3^-, which is transported across the proximal tubule (HCO_3^- regeneration). The NH_4^+ must be excreted in the urine for this process to be effective in bicarbonate regeneration. The excretion of ammonium involves secretion by proximal tubule cells (possibly by the Na^+/H^+ exchanger as Na^+/NH_4^+), generation of high medullary interstitial NH_4^+ concentration by an elaborate countercurrent multiplication/exchange system, and finally, secretion of the interstitial

NH_4^+ by the collecting duct by a combination of H^+ secretion and passive NH_3 diffusion. *Ammoniagenesis* is responsive to the acid-base needs of the individual. When faced with an acute acid burden and an increased need for HCO_3^- regeneration, the rate of renal ammonia synthesis increases sharply.

The quantity of hydrogen ions excreted as titratable acid and NH_4^+ is equal to the quantity of HCO_3^- regenerated in tubule cells and added to plasma. Under steady-state conditions, the net quantity of acid excreted into the urine (the sum of titratable acid and NH_4^+ less HCO_3^-) must equal the quantity of acid gained by the extracellular fluid from all sources. Metabolic acidosis and alkalosis result when this delicate balance is perturbed, the former the result of insufficient net acid excretion, and the latter due to excessive acid excretion.

Progressive loss of renal function usually causes little or no change in arterial pH, plasma bicarbonate concentration, or arterial carbon dioxide tension (P_{CO_2}) until GFR falls below 25% of normal. Thereafter, all three tend to decline as *metabolic acidosis* ensues. In general, the metabolic acidosis of CRF is not due to overproduction of acids but is rather a reflection of nephron loss, which limits the amount of NH_3 (and therefore also HCO_3^-) that can be generated. Although surviving nephrons appear to be capable of generating supranormal amounts of NH_3 *per nephron*, the diminished nephron population causes overall production to be reduced to an extent that is insufficient to permit adequate buffering of H^+ in urine. As a result, although patients with CRF may be able to acidify their urine normally (i.e., urine pH as low as 4.5), the defect in NH_3 production limits daily net acid excretion to 30 to 40 mmol, or one-half to two-thirds the quantity of nonvolatile acid added to the extracellular fluid in the same time period. Metabolic acidosis resulting from this daily positive balance of H^+ is seldom florid in CRF of mild to moderate severity. Relative stability of plasma bicarbonate (albeit at reduced levels of 14 to 18 mmol/L) is maintained at the expense of buffering by bone. Because it contains large reserves of alkaline salts (calcium phosphate and calcium bicarbonate), bone constitutes a major reserve of buffering capacity. Dissolution of these buffers contributes to the osteodystrophy of CRF (Fig. 270-1).

Although the acidosis of CRF is due to loss of tubule mass, it nevertheless depends to a large part on the level of GFR. When GFR is reduced to only a moderate extent (i.e., to about 50% of normal), retention of anions, principally sulfates and phosphates, is not pronounced. Therefore, as the plasma HCO_3^- falls owing to dysfunction or loss of tubules, retention of Cl^- by the kidneys leads to a *hyperchloremic acidosis*. At this stage *the anion gap is normal*. With further reductions in GFR and progressive azotemia, however, the retention of phosphates, sulfates, and other *unmeasured* anions ensues and plasma Cl^- falls to normal levels despite the reduction in plasma HCO_3^- concentration. *An elevated anion gap therefore develops.*

TUBULE POTASSIUM TRANSPORT WITH NORMAL AND REDUCED NEPHRON MASS

As with H^+, the concentration of K^+ in extracellular fluid is normally maintained within a relatively narrow range, 4 to 5 mmol/L. At least 95% of total-body K^+ is in the intracellular compartment, where the intracellular concentration is approximately 160 mmol/L. Normal individuals maintain external K^+ balance by excreting amounts into the

urine that equal the intake, less the relatively small losses in stool and sweat. K^+ is freely filtered at the glomerulus, although the amount excreted usually represents no more than about 20% of the quantity filtered. The great bulk of the K^+filtered is reabsorbed in the early portions of the nephron, about two-thirds in the proximal tubule, and an additional 20 to 25% in the loop of Henle. A *K^+secretory process* operates in the distal tubule and terminal nephron segments. This process is largely dependent on Na^+reabsorption and the accompanying lumen-negative voltage creating an electrical gradient across the tubule wall, favoring K^+secretion into the lumen of the distal tubule and collecting duct.

The ability to maintain external K^+balance and normal plasma K^+concentration until relatively late in the course ofCRF is a consequence primarily of a progressive increase in fractional excretion of K^+. Greatly enhanced rates of K^+secretion occur in distal portions of surviving tubules. The augmented secretion rate of aldosterone contributes to enhanced tubule secretion of K^+. In addition, both the increased distal tubule flow rates in surviving nephrons, due to the osmotic diuresis, and enhanced luminal electronegativity, created by the increased presence of highly impermeable anions such as phosphate and sulfate, enhance K^+secretion. Aldosterone also stimulates net entry of K^+ into the lumen of the colon, a mechanism known to be enhanced in CRF.**More detailed discussions of abnormal K^+homeostasis in acute and chronic forms of renal failure are given in Chaps. 269 and 270.*

ACKNOWLEDGEMENT

Harald S. MacKenzie was a co-author of thischapterin the 14th edition and some of the content of thatchapteris carried forward to the present edition.

(Bibliography omitted in Palm version)

Back to Table of Contents

269. ACUTE RENAL FAILURE - *Hugh R. Brady*, *Barry M. Brenner*

Acute renal failure (ARF) is a syndrome characterized by rapid decline in glomerular filtration rate (hours to days), retention of nitrogenous waste products, and perturbation of extracellular fluid volume and electrolyte and acid-base homeostasis. ARF complicates approximately 5% of hospital admissions and up to 30% of admissions to intensive care units. Oliguria (urine output< 500 mL/d) is a frequent but not invariable clinical feature (~50%). ARF is usually asymptomatic and is diagnosed when biochemical screening of hospitalized patients reveals a recent increase in plasma urea and creatinine concentrations. It may complicate a wide range of diseases, which for purposes of diagnosis and management are conveniently divided into three categories: (1) diseases that cause renal hypoperfusion without compromising the integrity of renal parenchyma (*prerenal ARF*, prerenal azotemia) (~55%); (2) diseases that directly involve renal parenchyma (*intrinsic renal ARF*, renal azotemia) (~40%); and (3) diseases associated with urinary tract obstruction (*postrenal ARF*, postrenal azotemia) (~5%). Most ARF is reversible, the kidney being relatively unique among major organs in its ability to recover from almost complete loss of function. Nevertheless, ARF is associated with major in-hospital morbidity and mortality, in large part due to the serious nature of the illnesses that precipitate the ARF.

ETIOLOGY AND PATHOPHYSIOLOGY

PRERENALARF(PRERENAL AZOTEMIA)

PrerenalARF is the most common form of ARF and represents a physiologic response to mild to moderate renal hypoperfusion. Prerenal ARF is rapidly reversible upon restoration of renal blood flow and glomerular ultrafiltration pressure. Renal parenchymal tissue is not damaged; indeed, kidneys from individuals with prerenal ARF function well when transplanted into recipients with normal cardiovascular function. More severe hypoperfusion may lead to ischemic injury of renal parenchyma and intrinsic renal ARF. Thus, prerenal ARF and intrinsic renal ARF due to ischemia are part of a spectrum of manifestations of renal hypoperfusion. As shown in Table 269-1, prerenal ARF can complicate any disease that induces hypovolemia, low cardiac output, systemic vasodilatation, or selective renal vasoconstriction.

Hypovolemia leads to a fall in mean systemic arterial pressure, which is detected as reduced stretch by arterial (e.g., carotid sinus) and cardiac baroreceptors. Activated baroreceptors trigger a coordinated series of neural and humoral responses designed to restore blood volume and arterial pressure. These include activation of the sympathetic nervous system and renin-angiotensin-aldosterone system and release of arginine vasopressin (AVP; formerly called antidiuretic hormone). Norepinephrine, angiotensin II, and AVP act in concert in an attempt to preserve cardiac and cerebral perfusion by stimulating vasoconstriction in relatively "nonessential" vascular beds, such as the musculocutaneous and splanchnic circulations, by inhibiting salt loss through sweat glands, by stimulating thirst and salt appetite, and by promoting renal salt and water retention. Glomerular perfusion, ultrafiltration pressure, and filtration rate are preserved during mild hypoperfusion through several compensatory mechanisms. Stretch receptors in afferent arterioles, in response to a reduction in perfusion pressure, trigger afferent arteriolar vasodilatation through a local myogenic reflex (autoregulation).

Biosynthesis of vasodilator prostaglandins (e.g., prostaglandin F_2 and prostacyclin) is also enhanced, and these compounds preferentially dilate afferent arterioles. In addition, angiotensin II induces preferential constriction of efferent arterioles. As a result, intraglomerular pressure is maintained, the fraction of plasma flowing through glomerular capillaries that is filtered is increased (filtration fraction), and glomerular filtration rate (GFR) is preserved. During states of more severe hypoperfusion, these compensatory responses are overwhelmed and GFR falls, leading to prerenalARF.

Autoregulatory dilatation of afferent arterioles is maximal at mean systemic arterial blood pressures of ~80 mmHg, and hypotension below this level is associated with a precipitous decline inGFR. Lesser degrees of hypotension may provoke prerenalARF in the elderly and in patients with diseases affecting the integrity of afferent arterioles (e.g., hypertensive nephrosclerosis, diabetic vasculopathy). In addition, drugs that interfere with adaptive responses in the renal microcirculation may convert compensated renal hypoperfusion into overt prerenal ARF or trigger progression of prerenal ARF to ischemic intrinsic renal ARF (see below). Pharmacologic inhibitors of either renal prostaglandin biosynthesis [*cyclooxygenase inhibitors*; nonsteroidal anti-inflammatory drugs (NSAIDs)] or angiotensin-converting enzyme (ACE) activity (ACE inhibitors) are the major culprits and should be used judiciously in the setting of suspected renal hypoperfusion. NSAIDs do not compromise GFR in healthy individuals but may precipitate prerenal ARF in patients with volume depletion or in those with chronic renal insufficiency in whom GFR is maintained, in part, through prostaglandin-mediated hyperfiltration through the remaining functional nephrons. ACE inhibitors can also compromise GFR in individuals with renal hypoperfusion and should be used with special care in patients with bilateral renal artery stenosis or unilateral stenosis in a solitary functioning kidney. Glomerular perfusion and filtration may be exquisitely dependent on the actions of angiotensin II under the latter circumstances. Angiotensin II preserves glomerular filtration pressure distal to stenoses by elevating systemic arterial pressure and by triggering selective constriction of efferent arterioles. ACE inhibitors blunt these responses and precipitate ARF, usually reversible, in ~30% of these patients.

Hepatorenal Syndrome This is a particularly aggressive form ofARF that frequently complicates hepatic failure due to advanced cirrhosis or other liver diseases, including malignancy, hepatic resection, and biliary obstruction. Intrarenal vasoconstriction and avid sodium retention are early sequelae of these diseases and may be detected before changes in systemic hemodynamics. Patients with advanced liver disease, portal hypertension, and ascites also have increased plasma volume but reduced "effective" arterial blood volume as a consequence of systemic vasodilatation and pooling of blood in the portal circulation. Renal failure typically develops slowly over weeks or months in parallel with deteriorating hepatic function but may accelerate dramatically following a variety of hemodynamic insults, including hemorrhage, paracentesis, and overzealous use of diuretics, vasodilators, or cyclooxygenase inhibitors. In full-blown hepatorenal syndrome, ARF progresses even after optimization of systemic hemodynamics and systemic arterial blood volume and removal of nephrotoxins, probably as a result of ongoing intrarenal vasoconstriction, hypoperfusion, and ischemia triggered by circulating factors or neural impulses originating in the failing liver. Indeed, it must be remembered that patients with liver disease may develop other forms of ARF (e.g., sepsis, nephrotoxic medications), and a diagnosis of hepatorenal syndrome should be

made only after exclusion of other possible reversible causes.

INTRINSIC RENAL ARF (INTRINSIC RENAL AZOTEMIA)

Intrinsic renal ARF can complicate many diverse diseases of the renal parenchyma. From a clinicopathologic viewpoint, it is useful to divide the causes of intrinsic renal ARF into (1) diseases of large renal vessels, (2) diseases of the renal microcirculation and glomeruli, (3) ischemic and nephrotoxic ARF, and (4) tubulointerstitial diseases (Table 269-1). Most intrinsic renal ARF is triggered by ischemia (ischemic ARF) or nephrotoxins (nephrotoxic ARF), insults that classically induce acute tubular necrosis (ATN). Accordingly, the terms ARF and ATN are usually used interchangeably in these settings. However, as many as 20 to 30% of patients with ischemic or nephrotoxic ARF do not have clinical (granular or tubular cell urinary casts) or morphologic evidence of tubular necrosis, underscoring the role of sublethal injury to tubular epithelium and injury to other renal cells (e.g., endothelial cells) in the pathophysiology of this syndrome.

Etiology and Pathophysiology of IschemicARF Prerenal ARF and ischemic ARF are part of a spectrum of manifestations of renal hypoperfusion. Ischemic ARF differs from prerenal ARF in that the hypoperfusion induces ischemic injury to renal parenchymal cells, particularly tubular epithelium, and recovery typically takes 1 to 2 weeks after normalization of renal perfusion as it requires repair and regeneration of renal cells. In its most extreme form, ischemia leads to bilateral renal cortical necrosis and irreversible renal failure. Ischemic ARF occurs most frequently in patients undergoing major cardiovascular surgery or suffering severe trauma, hemorrhage, sepsis, and/or volume depletion (Table 269-1). Ischemic ARF can also complicate milder forms of true hypovolemia or reduced "effective" arterial blood volume if they occur in the presence of other insults (e.g., nephrotoxins or sepsis) or in patients with compromised autoregulatory defense mechanisms or preexisting renal disease.

The course of ischemicARF is typically characterized by three phases: the initiation, maintenance, and recovery phases. The *initiation phase* (hours to days) is the initial period of renal hypoperfusion during which ischemic injury is evolving.GFRdeclines because (1) glomerular ultrafiltration pressure is reduced as a consequence of the fall in renal blood flow, (2) the flow of glomerular filtrate within tubules is obstructed by casts comprising epithelial cells and necrotic debris derived from ischemic tubule epithelium, and (3) there is backleak of glomerular filtrate through injured tubular epithelium (Fig. 269-1). Ischemic injury is most prominent in the terminal medullary portion of the proximal tubule (S_3segment, pars recta) and the medullary portion of the thick ascending limb of the loop of Henle. Both segments have high rates of active (ATP-dependent) solute transport and oxygen consumption and are located in a zone of the kidney (the outer medulla) that is relatively ischemic, even under basal conditions, by virtue of the unique countercurrent arrangement of the medullary vasculature. Cellular ischemia results in a series of alterations in energetics, ion transport, and membrane integrity that ultimately lead to cell injury and, if severe, cell apoptosis or necrosis. These alterations include depletion of ATP, inhibition of active sodium transport and transport of other solutes, impairment of cell volume regulation and cell swelling, cytoskeletal disruption and loss of cell polarity, cell-cell and cell-matrix attachment, accumulation of intracellular calcium, altered phospholipid metabolism, oxygen free radical formation, and peroxidation of membrane lipids. Importantly, renal

injury can be limited by restoration of renal blood flow during this period.

The initiation phase is followed by a *maintenance phase* (typically 1 to 2 weeks) during which renal cell injury is established,GFRstabilizes at its nadir (typically 5 to 10 mL/min), urine output is lowest, and uremic complications arise (see below). The reasons why the GFR remains low during this phase, despite correction of systemic hemodynamics, are still being defined. Putative mechanisms include persistent intrarenal vasoconstriction and medullary ischemia triggered by dysregulated release of vasoactive mediators from injured endothelial cells (e.g., decreased nitric oxide, increased endothelin-1 and platelet-activating factor), congestion of medullary blood vessels, and reperfusion injury induced by reactive oxygen species and other mediators derived from leukocytes or renal parenchymal cells (Fig. 269-1). In addition, epithelial cell injury per se may contribute to persistent intrarenal vasoconstriction by a process termed *tubuloglomerular feedback*. Specialized epithelial cells in the macula densa region of distal tubules detect increases in distal salt (probably chloride) delivery that occur as a consequence of impaired reabsorption by more proximal nephron segments. Macula densa cells in turn stimulate constriction of adjacent afferent arterioles by a poorly defined mechanism and further compromise glomerular perfusion and filtration, thereby contributing to a vicious cycle. A *recovery phase* is characterized by renal parenchymal cell, particularly tubule epithelial cell, repair and regeneration and a gradual return of GFR to or towards premorbid levels. The recovery phase may be complicated by a marked diuretic phase due to excretion of retained salt and water and other solutes, continued use of diuretics, and/or delayed recovery of epithelial cell function (solute and water reabsorption) relative to glomerular filtration (see below).

Etiology and Pathophysiology of Nephrotoxic**ARF**Acute intrinsic renal ARF can complicate exposure to many structurally diverse pharmacologic agents (Table 269-1). With most nephrotoxins, the incidence of ARF is increased in the elderly and in patients with preexisting chronic renal insufficiency, true or "effective" hypovolemia, or concomitant exposure to other toxins.

Intrarenal vasoconstriction is a pivotal event inARFtriggered by *radiocontrast agents* (contrast nephropathy) and *cyclosporine*. In keeping with this pathophysiology, both agents induce ARF that shares features with prerenal ARF: namely, an acute fall in renal blood flow andGFR, a relatively benign urine sediment, and a low fractional excretion of sodium (see below). Severe cases may show clinical or pathologic evidence ofATN. Contrast nephropathy classically presents as an acute (onset within 24 to 48 h) but reversible (peak 3 to 5 days, resolution within 1 week) rise in blood urea nitrogen and creatinine and is most common in individuals with preexisting chronic renal insufficiency, diabetes mellitus, congestive heart failure, hypovolemia, or multiple myeloma. The syndrome appears to be dose-related, and its incidence is only slightly reduced in high-risk individuals by use of more expensive low osmolality, nonionic contrast agents. Endothelin-1, a potent vasoconstrictor peptide released from endothelial cells, is an important mediator of intrarenal vasoconstriction and mesangial cell contraction in this setting. Endothelin-1 has also been implicated as an important mediator of cyclosporine-induced ARF.

Direct toxicity to tubule epithelial cells and/or intratubular obstruction are major pathophysiologic events inARFinduced by many antibiotics and anticancer drugs.

Frequent offenders are the antimicrobial agents, such as acyclovir, foscarnet, aminoglycosides, amphotericin B, and pentamidine, and chemotherapeutic agents, such as cisplatin and ifosfamide. ARF complicates 10 to 30% of courses of *aminoglycoside antibiotics*, even in the presence of therapeutic levels. *Amphotericin B* causes dose-related ARF through intrarenal vasoconstriction and direct toxicity to proximal tubule epithelium. Cisplatin, like the aminoglycosides, is accumulated by proximal tubule cells and typically provokes ARF after 7 to 10 days of exposure by inducing mitochondrial injury, inhibition of ATPase activity and solute transport, free radical-mediated injury to cell membranes, apoptosis and/or necrosis.

The most common endogenous nephrotoxins are calcium, myoglobin, hemoglobin, urate, oxalate, and myeloma light chains. Hypercalcemia can compromiseGFR, predominantly by inducing intrarenal vasoconstriction. Calcium phosphate deposition within the kidney may also contribute. Both *rhabdomyolysis* and *hemolysis* can induceARF, particularly in hypovolemic or acidotic individuals. Myoglobinuric ARF complicates approximately 30% of cases of rhabdomyolysis. Common causes of the latter include traumatic crush injury, acute muscle ischemia, seizures, excessive exercise, heat stroke or malignant hyperthermia, intoxications (e.g., alcohol, cocaine), and infectious or metabolic disorders. ARF due to hemolysis is relatively rare and is observed following massive blood transfusion reactions. It has been postulated that myoglobin and hemoglobin or other compounds released from muscle or red blood cells cause ARF via toxic effects on tubule epithelial cells or by inducing intratubular cast formation. Hypovolemia or acidosis may contribute to the pathogenesis of ARF in this setting by promoting intratubular cast formation. In addition, both hemoglobin and myoglobin are potent inhibitors of nitric oxide bioactivity and may trigger intrarenal vasoconstriction and ischemia in patients with borderline renal hypoperfusion. The formation of intratubular casts containing filtered immunoglobulin light chains and other proteins, including Tamm-Horsfall protein produced by thick ascending limb cells, is the major trigger for ARF in patients with *multiple myeloma* (myeloma cast nephropathy). Light chains may also be directly toxic to tubule epithelial cells. Intratubular obstruction may also be an important cause of ARF in patients with severe *hyperuricosuria* or *hyperoxaluria*. Acute uric acid nephropathy typically complicates treatment of lymphoproliferative or myeloproliferative disorders but occasionally occurs in other forms of primary or secondary hyperuricemia if the urine is concentrated.

Pathology of Ischemic and Nephrotoxic ARF The classic pathologic features of ischemic ARF are patchy and focal necrosis of tubule epithelium with detachment from its basement membrane and occlusion of tubule lumens with casts composed of intact or degenerating epithelial cells, cellular debris, Tamm-Horsfall mucoprotein, and pigments. Leukocyte accumulation is frequently observed in vasa recta; however, the morphology of the glomeruli and renal vasculature is characteristically normal. Necrosis is most severe in the straight portion (pars recta) of proximal tubules but may also affect the medullary thick ascending limb of the loop of Henle.

In nephrotoxicARF, morphologic changes tend to be most prominent in both the convoluted and straight portions of proximal tubules. Tubule cell necrosis is less pronounced than in ischemic ARF.

Other Causes of Intrinsic Renal ARF Patients with advanced atherosclerosis can

develop ARF after manipulation of the aorta or renal arteries at surgery or angiography, following trauma, or, rarely, spontaneously due to embolization of cholesterol crystals to the renal vasculature (atheroembolic ARF). Cholesterol crystals lodge in small- and medium-sized arteries and incite a giant cell and fibrotic reaction in the vessel wall with narrowing or obstruction of the vessel lumen. Atheroembolic ARF is frequently irreversible. A myriad of structurally diverse pharmacologic agents induce ARF by triggering allergic interstitial nephritis, a disease characterized by infiltration of the tubulointerstitium by granulocytes (typically but not invariably eosinophils), macrophages, and/or lymphocytes and by interstitial edema. The most common offenders are antibiotics (e.g., penicillins, cephalosporins, trimethoprim, sulfonamides, rifampicin) and NSAIDs (Table 269-1).

POSTRENALARF(See also Chap. 281)

Urinary tract obstruction accounts for fewer than 5% of cases of ARF. Since one kidney has sufficient clearance capacity to excrete nitrogenous waste products, ARF from obstruction requires either obstruction to urine flow between the external urethral meatus and bladder neck, bilateral ureteric obstruction, or unilateral ureteric obstruction in a patient with one functioning kidney or preexisting chronic renal insufficiency. Bladder neck obstruction represents the most common cause of postrenal ARF and is usually due to prostatic disease (e.g., hypertrophy, neoplasia, or infection), neurogenic bladder, or therapy with anticholinergic drugs. Less common causes of acute lower urinary tract obstruction include blood clots, calculi, and urethritis with spasm. Ureteric obstruction may result from intraluminal obstruction (e.g., calculi, blood clots, sloughed renal papillae), infiltration of the ureteric wall (e.g., neoplasia), or external compression (e.g., retroperitoneal fibrosis, neoplasia or abscess, inadvertent surgical ligature). During the early stages of obstruction (hours to days), continued glomerular filtration leads to increased intraluminal pressure upstream to the site of obstruction. As a result there are gradual distention of proximal ureter, renal pelvis, and calyces and a fall inGFR. Acute obstruction is initially associated with modest increase in renal blood flow, but arteriolar vasoconstriction soon supervenes, leading to a further decline in glomerular filtration.

CLINICAL FEATURES AND DIFFERENTIAL DIAGNOSIS

Patients presenting with renal failure should be assessed initially to determine if the decline inGFR is acute or chronic. An acute process is easily established if a review of laboratory records reveals a recent rise in blood urea and creatinine levels, but previous measurements are not always available. Findings that suggest chronic renal failure (Chap. 270) include anemia, neuropathy, and radiologic evidence of renal osteodystrophy or small scarred kidneys. However, it should be noted that anemia may also complicateARF (see below), and renal size may be normal or increased in several chronic renal diseases (e.g., diabetic nephropathy, amyloidosis, polycystic kidney disease). Once a diagnosis of ARF has been established, several issues should be addressed promptly: (1) the identification of the cause of ARF, (2) the elimination of the triggering insult (e.g., nephrotoxin) and/or institution of disease-specific therapies, and (3) the prevention and management of uremic complications.

CLINICAL ASSESSMENT

Clinical clues to *prerenal*ARF are symptoms of thirst and orthostatic dizziness and physical evidence of orthostatic hypotension and tachycardia, reduced jugular venous pressure, decreased skin turgor, dry mucous membranes, and reduced axillary sweating. Case records should be reviewed for documentation of a progressive fall in urine output and body weight and treatment with NSAIDs orACEinhibitors. Careful clinical examination may reveal stigmata of chronic liver disease and portal hypertension, advanced cardiac failure, sepsis, or other causes of reduced "effective" arterial blood volume (Table 269-1).

*Intrinsic renal*ARF due to ischemia is likely following severe renal hypoperfusion complicating hypovolemic or septic shock or following major surgery. The likelihood of ischemic ARF is increased further if ARF persists despite normalization of systemic hemodynamics. Diagnosis of nephrotoxic ARF requires careful review of the clinical data and pharmacy, nursing, and radiology records for evidence of recent exposure to nephrotoxic medications or radiocontrast agents or to endogenous toxins (e.g., myoglobin, hemoglobin, uric acid, myeloma protein, or elevated levels of serum calcium).

Although ischemic and nephrotoxicARFaccount for more than 90% of cases of intrinsic renal ARF, other renal parenchymal diseases must be considered (Table 269-2). Flank pain may be a prominent symptom following occlusion of a renal artery or vein and with other parenchymal diseases distending the renal capsule (e.g., severe glomerulonephritis or pyelonephritis). Subcutaneous nodules, livido reticularis, bright orange retinal arteriolar plaques, and digital ischemia, despite palpable pedal pulses, are clues to atheroembolization. ARF in association with oliguria, edema, hypertension, and an "active" urine sediment (nephritic syndrome) suggests acute glomerulonephritis or vasculitis. Malignant hypertension is a likely cause of ARF in patients with severe hypertension and evidence of hypertensive injury to other organs (e.g., left ventricular hypertrophy and failure, hypertensive retinopathy and papilledema, neurologic dysfunction). Fever, arthralgias, and a pruritic erythematous rash following exposure to a new drug suggest allergic interstitial nephritis, although systemic features of hypersensitivity are frequently absent.

*Postrenal*ARFpresents with suprapubic and flank pain due to distention of the bladder and of the renal collecting system and capsule, respectively. Colicky flank pain radiating to the groin suggests acute ureteric obstruction. Prostatic disease is likely if there is a history of nocturia, frequency, and hesitancy and enlargement or induration of the prostate on rectal examination. Neurogenic bladder should be suspected in patients receiving anticholinergic medications or with physical evidence of autonomic dysfunction. Definitive diagnosis of postrenal ARF hinges on judicious use of radiologic investigations and rapid improvement in renal function following relief of obstruction.

URINALYSIS

Anuria suggests complete urinary tract obstruction but may complicate severe cases of prerenal or intrinsic renal ARF. Wide fluctuations in urine output raise the possibility of intermittent obstruction, whereas patients with partial urinary tract obstruction can present with polyuria due to impairment of urine concentrating mechanisms.

In prerenalARF, the sediment is characteristically acellular and contains transparent hyaline casts ("bland," "benign," "inactive" urine sediment). Hyaline casts are formed in concentrated urine from normal constitutents of urine -- principally Tamm-Horsfall protein, which is secreted by epithelial cells of the loop of Henle. Postrenal ARF may also present with an inactive sediment, although hematuria and pyuria are common in patients with intraluminal obstruction or prostatic disease. Pigmented "muddy brown" granular casts and casts containing tubule epithelial cells are characteristic ofATN and suggest ischemic or nephrotoxic ARF. They are usually found in association with microscopic hematuria and mild "tubular" proteinuria (<1 g/d); the latter reflects impaired reabsorption and processing of filtered proteins by injured proximal tubules. Casts are absent, however, in 20 to 30% of patients with ischemic or nephrotoxic ARF and are not a requisite for diagnosis. In general, red blood cell casts indicate glomerular injury or, less often, acute tubulointerstitial nephritis. White cell casts and nonpigmented granular casts suggest interstitial nephritis, whereas broad granular casts are characteristic of chronic renal disease and probably reflect interstitial fibrosis and dilatation of tubules. Eosinophiluria (>5% of urine leukocytes) is a common finding (~90%) in antibiotic-induced allergic interstitial nephritis when studied using Hansel's stain; however, lymphocytes may predominate in allergic interstitial nephritis induced by NSAIDs. Eosinophiluria is also a feature of atheroembolic ARF. Occasional uric acid crystals (pleomorphic in shape) are common in the concentrated urine of prerenal ARF but suggest acute urate nephropathy if seen in abundance. Oxalate (envelope-shaped) and hippurate (needle-shaped) crystals raise the possibility of ethylene glycol ingestion and toxicity.

Increased urine protein excretion, but <1 g/d, is common inATN due to failure of injured proximal tubules to reabsorb filtered protein and excretion of cellular debris ("tubular proteinuria"). Proteinuria of>1 g/d suggests injury to the glomerular ultrafiltration barrier ("glomerular proteinuria") or excretion of myeloma light chains. The latter are not detected by conventional dipsticks (which detect albumin) and must be sought by other means (e.g., sulfosalicylic acid test, immunoelectrophoresis). Heavy proteinuria is also a frequent finding (~80%) in patients who develop combined allergic interstitial nephritis and minimal change glomerulopathy when treated with NSAIDs. A similar syndrome can be triggered by ampicillin, rifampicin, or interferon a. Hemoglobinuria or myoglobinuria should be suspected if urine is strongly positive for heme by dipstick, but contains few red cells, and if the supernatant of centrifuged urine is positive for free heme. Bilirubinuria may provide a clue to the presence of hepatorenal syndrome.

RENAL FAILURE INDICES

Analysis of urine and blood biochemistry is particularly useful for distinguishing prerenalARF from ischemic or nephrotoxic intrinsic renal ARF (Table 269-3). The fractional excretion of sodium (FE_{Na}) is most useful in this regard. The FE_{Na}relates sodium clearance to creatinine clearance. Sodium is reabsorbed avidly from glomerular filtrate in patients with prerenal ARF, in an attempt to restore intravascular volume, but not in patients with ischemic or nephrotoxic intrinsic ARF, as a result of tubular epithelial cell injury. In contrast, creatinine is not reabsorbed in either setting. Consequently, patients with prerenal ARF typically have a FE_{Na} of<1.0% (frequently<0.1%), whereas the FE_{Na} in patients with ischemic or nephrotoxic ARF is usually>1.0%. The *renal failure index* (Table 269-3) provides comparable information, since clinical variations in serum

sodium concentration are relatively small. *Urine sodium concentration* is a less sensitive index for distinguishing prerenal ARF from ischemic and nephrotoxic ARF as values overlap between groups. Similarly, indices of urinary concentrating ability such as urine specific gravity, urine osmolality, urine-to-plasma urea ratio, and blood urea-to-creatinine ratio are of limited value in differential diagnosis.

Many caveats apply when interpreting biochemical renal failure indices. FE_{Na} may be >1.0% in prerenal ARF if patients are receiving diuretics or have bicarbonaturia (accompanied by sodium to maintain electroneutrality), preexisting chronic renal failure complicated by salt wasting, or adrenal insufficiency. In contrast, the FE_{Na} is<1.0% in approximately 15% of patients with nonoliguric ischemic or nephrotoxic ARF. The FE_{Na} is often <1.0% in ARF due to urinary tract obstruction, glomerulonephritis, and vascular diseases.

LABORATORY FINDINGS

Serial measurements of serum creatinine can provide useful pointers to the cause of ARF. Prerenal ARF is typified by fluctuating levels that parallel changes in hemodynamic function. Creatinine rises rapidly (within 24 to 48 h) in patients with ARF following renal ischemia, atheroembolization, and radiocontrast exposure. Peak creatinine levels are observed after 3 to 5 days with contrast nephropathy and return to baseline after 5 to 7 days. In contrast, creatinine levels typically peak later (7 to 10 days) in ischemic ARF and atheroembolic disease. The initial rise in serum creatinine is characteristically delayed until the second week of therapy with many tubule epithelial cell toxins (e.g., aminoglycosides, cisplatin) and probably reflects the need for accumulation of these agents within cells before GFR falls.

Hyperkalemia, hyperphosphatemia, hypocalcemia, and elevations in serum uric acid and creatine kinase (MM isoenzyme) levels at presentation suggest a diagnosis of rhabdomyolysis. Hyperuricemia [>890 umol/L (>15 mg/dL)] in association with hyperkalemia, hyperphosphatemia, and increased circulating levels of intracellular enzymes such as lactate dehydrogenase may indicate acute urate nephropathy and tumor lysis syndrome following cancer chemotherapy. A wide serum anion and osmolal gap (measured serum osmolality minus the serum osmolality calculated from serum sodium, glucose, and urea concentrations) indicate the presence of an unusual anion or osmole in the circulation and are clues to diagnosis of ethylene glycol or methanol ingestion. Severe anemia in the absence of hemorrhage raises the possibility of hemolysis, multiple myeloma, or thrombotic microangiopathy. Systemic eosinophilia suggests allergic interstitial nephritis but is also a feature of atheroembolic disease and polyangiitis nodosa.

RADIOLOGIC FINDINGS

Imaging of the urinary tract by ultrasonography is useful to exclude postrenal ARF. Computed tomography and magnetic resonance imaging are alternative imaging modalities. Whereas pelvicalyceal dilatation is usual with urinary tract obstruction (98% sensitivity), dilatation may be absent immediately following obstruction or in patients with ureteric encasement (e.g., retroperitoneal fibrosis, neoplasia). Retrograde or anterograde pyelography are more definitive investigations in complex cases and

provide precise localization of the site of obstruction. A plain film of the abdomen, with tomography if necessary, is a valuable initial screening technique in patients with suspected nephrolithiasis. Doppler ultrasonography and magnetic resonance flow imaging appear promising for assessment of patency of renal arteries and veins in patients with suspected vascular obstruction; however, contrast angiography is usually required for definitive diagnosis.

RENAL BIOPSY

Biopsy is reserved for patients in whom prerenal and postrenalARF have been excluded and the cause of intrinsic renal ARF is unclear. Renal biopsy is particularly useful when clinical assessment and laboratory investigations suggest diagnoses other than ischemic or nephrotoxic injury that may respond to disease-specific therapy. Examples include glomerulonephritis, vasculitis, hemolytic-uremic syndrome, thrombotic thrombocytopenic purpura, and allergic interstitial nephritis.

COMPLICATIONS

ARFimpairs renal excretion of sodium, potassium, and water and perturbs divalent cation homeostasis and urinary acidification mechanisms. As a result, ARF is frequently complicated by intravascular volume overload, hyponatremia, hyperkalemia, hyperphosphatemia, hypocalcemia, hypermagnesemia, and metabolic acidosis. In addition, patients are unable to excrete nitrogenous waste products and are prone to develop the uremic syndrome (Chap. 270). The speed of development and the severity of these complications reflect the degree of renal impairment and catabolic state of the patient.

Expansion of extracellular fluid volume is an inevitable consequence of diminished salt and water excretion in oliguric or anuric individuals. Whereas milder forms are characterized by weight gain, bibasilar lung rales, raised jugular venous pressure, and dependent edema, continued volume expansion may precipitate life-threatening pulmonary edema. Hypervolemia may be particularly problematic in patients receiving multiple intravenous medications and enteral or parenteral nutrition. Excessive administration of free water either through ingestion and nasogastric administration or as hypotonic saline or isotonic dextrose solutions (dextrose being metabolized) can induce *hypoosmolality* and *hyponatremia*, which, if severe, lead to cerebral edema and neurologic abnormalities, including seizures.

Hyperkalemia is a frequent complication ofARF. Serum potassium typically rises by 0.5 mmol/L per day in oliguric and anuric patients due to impaired excretion of ingested or infused potassium and potassium released from injured tissue. Coexistent metabolic acidosis may exacerbate hyperkalemia by promoting potassium efflux from cells. Hyperkalemia may be particularly severe, even at the time of diagnosis, in patients with rhabdomyolysis, hemolysis, and tumor lysis syndrome. Mild hyperkalemia (<6.0 mmol/L) is usually asymptomatic. Higher levels are typically associated with electrocardiographic abnormalities and/or increased cardiac excitability (Chap. 226).

Metabolism of dietary protein yields between 50 and 100 mmol/d of fixed nonvolatile acids that are normally excreted by the kidneys. Consequently,ARF is typically

complicated by *metabolic acidosis*, often with an increased serum anion gap (Chap. 50). Acidosis can be particularly severe when endogenous production of hydrogen ions is increased by other mechanisms (e.g., diabetic or fasting ketoacidosis; lactic acidosis complicating generalized tissue hypoperfusion, liver disease, or sepsis; metabolism of ethylene glycol or methanol).

Mild *hyperphosphatemia* is an almost invariable complication ofARF. Severe hyperphosphatemia may develop in highly catabolic patients or following rhabdomyolysis, hemolysis, or tumor lysis. Metastatic deposition of calcium phosphate can lead to *hypocalcemia*, particularly when the product of serum calcium (mg/dL) and phosphate (mg/dL) concentrations exceeds 70. Other factors that contribute to hypocalcemia include tissue resistance to the actions of parathyroid hormone and reduced levels of 1,25-dihydroxyvitamin D. Hypocalcemia is often asymptomatic but can cause perioral paresthesias, muscle cramps, seizures, hallucinations and confusion, and prolongation of the QT interval and nonspecific T-wave changes on electrocardiography (Chap. 341).

Anemia develops rapidly inARF and is usually mild and multifactorial in origin. Contributing factors include impaired erythropoiesis, hemolysis, bleeding, hemodilution, and reduced red cell survival time. Prolongation of the *bleeding time* and *leukocytosis* are also common. Common contributors to the bleeding diathesis include mild thrombocytopenia, platelet dysfunction, and/or clotting factor abnormalities (e.g., factor VIII dysfunction), whereas leukocytosis usually reflects sepsis, a stress response, or other concurrent illness. *Infection* is a common and serious complication of ARF, occurring in 50 to 90% of cases and accounting for up to 75% of deaths. It is unclear whether patients with ARF have a clinically significant defect in host immune responses or whether the high incidence of infection reflects repeated breaches of mucocutaneous barriers (e.g., intravenous cannulae, mechanical ventilation, bladder catheterization). *Cardiopulmonary complications* of ARF include arrhythmias, myocardial infarction, pericarditis and pericardial effusion, pulmonary edema, and pulmonary embolism. Mild *gastrointestinal bleeding* is common (10 to 30%) and is usually due to stress ulceration of gastric or small intestinal mucosa.

Protracted periods of severeARF are invariably associated with the development of the *uremic syndrome* (Chap. 270).

A *vigorous diuresis* can occur during the recovery phase ofARF (see above) and lead to intravascular volume depletion and delayed recovery ofGFR by causing secondary prerenal ARF. *Hypernatremia* can also complicate recovery if water losses via hypotonic urine are not replaced or if losses are inappropriately replaced by relatively hypertonic saline solutions. *Hypokalemia*, *hypomagnesemia*, *hypophosphatemia*, and *hypocalcemia* are less common metabolic complications during this period.

TREATMENT

Prevention Because there are no specific therapies for ischemic or nephrotoxicARF, prevention is of paramount importance. Many cases of ischemic ARF can be avoided by close attention to cardiovascular function and intravascular volume in high-risk patients, such as the elderly and those with preexisting renal insufficiency. Indeed, aggressive

restoration of intravascular volume has been shown to reduce the incidence of ischemic ARF dramatically after major surgery or trauma, burns, or cholera. The incidence of nephrotoxic ARF can be reduced by tailoring the dosage of potential nephrotoxins to body size andGFR; for example, reducing the dose or frequency of administration of drugs in patients with preexisting renal impairment. In this regard, it should be noted that serum creatinine is a relatively insensitive index of GFR and may overestimate GFR considerably in small or elderly patients. For purposes of drug dosing, it is advisable to estimate the GFR using the Cockcroft-Gault formula, which factors in the variables of age and weight (Chap. 47). Adjusting drug dosage according to circulating drug levels also appears to limit renal injury in patients receiving aminoglycoside antibiotics or cyclosporine. Diuretics, cyclooxygenase inhibitors,ACEinhibitors, and other vasodilators should be used with caution in patients with suspected true or "effective" hypovolemia or renovascular disease as they may precipitate prerenal ARF or convert the latter to ischemic ARF. Hypovolemia should be avoided in patients receiving nephrotoxic medications as renal hypoperfusion potentiates the toxicity of most nephrotoxins. Allopurinol and forced alkaline diuresis are useful in patients at high risk for acute urate nephropathy (e.g., cancer chemotherapy in hematologic malignancies) to limit uric acid generation and prevent precipitation of urate crystals in renal tubules. Forced alkaline diuresis may also prevent or attenuate ARF in patients receiving high-dose methotrexate or suffering from rhabdomyolysis. *N*-acetylcysteine limits acetaminophen-induced renal injury if given within 24 h of ingestion. Dimercaprol, a chelating agent, may prevent heavy metal nephrotoxicity. Ethanol inhibits ethylene glycol metabolism to oxalic acid and other toxic metabolites and is an important adjunct to hemodialysis in the emergency management of ethylene glycol intoxication.

Specific Therapies By definition, prerenalARF is rapidly reversible upon correction of the primary hemodynamic abnormality, and postrenal ARF resolves upon relief of obstruction. To date, there are no specific therapies for established intrinsic renal ARF due to ischemia or nephrotoxicity. Management of these disorders should focus on elimination of the causative hemodynamic abnormality or toxin, avoidance of additional insults, and prevention and treatment of complications. Specific treatment of other causes of intrinsic renal ARF depends on the underlying pathology.

Prerenal ARF The composition of replacement fluids for treatment of prerenal ARF due to hypovolemia must be tailored according to the composition of the lost fluid. Severe hypovolemia due to hemorrhage should be corrected with packed red blood cells, whereas isotonic saline is usually appropriate replacement for mild to moderate hemorrhage or plasma loss (e.g., burns, pancreatitis). Urinary and gastrointestinal fluids can vary greatly in composition but are usually hypotonic. Hypotonic solutions (e.g., 0.45% saline) are usually recommended as initial replacement in patients with prerenal ARF due to increased urinary or gastrointestinal fluid losses, although isotonic saline may be more appropriate in severe cases. Subsequent therapy should be based on measurements of the volume and ionic content of excreted or drained fluids. Serum potassium and acid-base status should be monitored carefully, and potassium and bicarbonate supplemented as appropriate. Cardiac failure may require aggressive management with positive inotropes, preload and afterload reducing agents, antiarrhythmic drugs, and mechanical aids such as intraaortic balloon pumps. Invasive hemodynamic monitoring may be required to guide therapy for complicated conditions in patients in whom clinical assessment of cardiovascular function and intravascular

volume proves unreliable.

Fluid management may be particularly difficult in patients with cirrhosis complicated by ascites. In this setting, it is important to distinguish between full-blown hepatorenal syndrome (Chap. 299), which carries a grave prognosis, and reversibleARF due to true or "effective" hypovolemia induced by overzealous use of diuretics or sepsis (e.g., spontaneous bacterial peritonitis). The contribution of hypovolemia to ARF can be definitively assessed only by administration of a fluid challenge. Fluids should be administered slowly and titrated against jugular venous pressure and, if necessary, central venous and pulmonary capillary wedge pressure, abdominal girth, and urine output. Patients with a reversible prerenal component typically have an increase in urine output and fall in serum creatinine, whereas patients with hepatorenal syndrome do not and may suffer increased ascites formation and pulmonary compromise if not monitored closely. Large volumes of ascitic fluid can usually be drained by paracentesis without deterioration in renal function if intravenous albumin is administered simultaneously. Indeed, "large-volume paracentesis" may afford an increase inGFR, possibly by lowering intraabdominal pressure and improving flow in renal veins. Shunting of ascitic fluid from the peritoneum to a central vein (peritoneojugular shunt, LeVeen or Denver shunts) is an alternative approach in refractory cases but has not been shown to improve survival in controlled trials. The efficacy of the newer technique of transjugular intrahepatic portosystemic shunting (TIPS procedure) is currently undergoing rigorous clinical assessment. Shunting can also improve GFR and sodium excretion transiently, probably because the increase in central blood volume stimulates release of atrial natriuretic peptides and inhibits secretion of aldosterone and norepinephrine.

Intrinsic renal ARF Many different approaches have been tested for their ability to attenuate injury or hasten recovery in ischemic and nephrotoxic ARF. These include atrial natriuretic peptide (ANP), low-dose dopamine, loop-blocking diuretics, calcium channel blockers, a-adrenoreceptor blockers, prostaglandin analogues, antioxidants, antibodies against leukocyte adhesion molecules, and insulin-like growth factor. Whereas many of these are beneficial in experimental models of ischemic or nephrotoxic ARF, they have either failed to confer consistent benefit or proved ineffective in humans.

ARFdue to other intrinsic renal diseases such as acute glomerulonephritis or vasculitis may respond to glucocorticoids, alkylating agents, and/or plasmapheresis, depending on the primary pathology. Glucocorticoids also hasten remission in some cases of allergic interstitial nephritis. Aggressive control of systemic arterial pressure is of paramount importance in limiting renal injury in malignant hypertensive nephrosclerosis, toxemia of pregnancy, and other vascular diseases. Hypertension and ARF due to scleroderma may be exquisitely sensitive to treatment withACEinhibitors.

Postrenal ARF Management of postrenal ARF requires close collaboration between nephrologist, urologist, and radiologist. Obstruction of the urethra or bladder neck is usually managed initially by transurethral or suprapubic placement of a bladder catheter, which provides temporary relief while the obstructing lesion is identified and treated definitively. Similarly, ureteric obstruction may be treated initially by percutaneous catheterization of the dilated renal pelvis or ureter. Indeed, obstructing lesions can often be removed percutaneously (e.g., calculus, sloughed papilla) or bypassed by insertion

of a ureteric stent (e.g., carcinoma). Most patients experience an appropriate diuresis for several days following relief of obstruction. Approximately 5% of patients develop a transient salt-wasting syndrome that may require administration of intravenous saline to maintain blood pressure.

Supportive Measures (Table 269-4) Following correction of hypovolemia, salt and water intake are tailored to match losses. Hypervolemia can usually be managed by restriction of salt and water intake and diuretics. Indeed, there is, as yet, no proven rationale for administration of diuretics inARFexcept to treat this complication. High doses of loop-blocking diuretics such as furosemide (up to 200 to 400 mg intravenously) or bumetanide (up to 10 mg intravenously administered as a bolus or by continuous infusion) may promote diuresis in patients who fail to respond to conventional doses. Subpressor doses of dopamine are claimed to promote salt and water excretion by increasing renal blood flow andGFR and by inhibiting tubule sodium reabsorption; however, subpressor ("low-dose," "renal-dose,") dopamine has proved ineffective in clinical trials and may trigger arrythmias and sudden cardiac death in critically ill patients. Ultrafiltration or dialysis is used to treat severe hypervolemia when conservative measures fail. Hyponatremia and hypoosmolality can usually be controlled by restriction of free water intake. Conversely, hypernatremia is treated by administration of water or intravenous hypotonic saline or isotonic dextrose-containing solutions.

**The management of hyperkalemia is described in Chap. 49.*

Metabolic acidosis is not treated unless serum bicarbonate concentration falls below 15 mmol/L or arterial pH falls below 7.2. More severe acidosis is corrected by oral or intravenous sodium bicarbonate. Initial rates of replacement are guided by estimates of bicarbonate deficit and adjusted thereafter according to serum levels (Chap. 50). Patients are monitored for complications of bicarbonate administration such as hypervolemia, metabolic alkalosis, hypocalcemia, and hypokalemia. From a practical point of view, most patients requiring sodium bicarbonate need emergency dialysis within days. Hyperphosphatemia is usually controlled by restriction of dietary phosphate and by oral aluminum hydroxide or calcium carbonate, which reduce gastrointestinal absorption of phosphate. Hypocalcemia does not usually require treatment unless severe, as may occur with rhabdomyolysis or pancreatitis or following admininstration of bicarbonate. Hyperuricemia is typically mild [<890 umol/L (< 15 mg/dL)] and does not require intervention.

The objective of *nutritional management* during the maintenance phase ofARF is to provide sufficient calories to avoid catabolism and starvation ketoacidosis, while minimizing production of nitrogenous waste. This is best achieved by restricting dietary protein to approximately 0.6 g/kg per day of protein of high biologic value (i.e., rich in essential amino acids) and to provide most calories as carbohydrate (approximately 100 g daily). Nutritional management is easier in nonoliguric patients and following institution of dialysis. Vigorous parenteral hyperalimentation is claimed to improve prognosis; however, convincing benefit has yet to be demonstrated in controlled trials.

Anemia may necessitate blood transfusion if severe or if recovery is delayed. In contrast to chronic renal failure, recombinant human erythropoietin is rarely used inARFbecause

bone marrow resistance to erythropoietin is common, more immediate treatment of anemia (if any) is required, and renal failure is usually self-limiting. Uremic bleeding usually responds to correction of anemia, administration of desmopressin or estrogens, or dialysis. Regular doses of antacids appear to reduce the incidence of gastrointestinal hemorrhage significantly and may be more effective in this regard than H_2antagonists, or proton pump inhibitors. Meticulous care of intravenous cannulae, bladder catheters, and other invasive devices is mandatory to avoid infections. Unfortunately, prophylactic antibiotics have not been shown to reduce the incidence of infection in these high-risk patients.

Indications and Modalities of Dialysis Dialysis replaces renal function until regeneration and repair restore renal function. Hemodialysis and peritoneal dialysis appear equally effective for management ofARF. Thus, the dialysis modality is chosen according to the needs of individual patients (e.g., peritoneal dialysis may be preferable if the patient is hemodynamically unstable, and hemodialysis after abdominal surgery involving the peritoneum), the expertise of the nephrologist, and the facilities of the institution. Vascular access for conventional intermittent hemodialysis is best achieved by insertion of a temporary double-lumen hemodialysis catheter into the internal jugular vein. The subclavian and femoral veins are alternative access sites. Peritoneal dialysis is achieved by insertion of a single-lumen cuffed catheter into the peritoneal cavity. Absolute indications for dialysis include symptoms or signs of the uremic syndrome and management of refractory hypervolemia, hyperkalemia, or acidosis. Many nephrologists also initiate dialysis empirically for blood urea levels of >100 mg/dL, even in the absence of clinical uremia; however, this approach has yet to be validated in controlled clinical trials. Nor is it clear whether intensive dialysis prescribed to maintain blood urea and creatinine below a certain level is beneficial. The latter are important issues since unnecessary or intensive hemodialysis can exacerbateATN and delay renal recovery by triggering hypotension and repeated renal hypoperfusion. Moreover, an expanding body of evidence suggests that leukocytes, activated directly by contact with hemodialysis membranes or as a result of membrane-triggered complement activation, then travel to the already-compromised renal microcirculation where they further exacerbate renal injury.

Continuous arteriovenous hemodiafiltration (CAVH) and continuous venovenous hemodiafiltration (CVVH) are alternatives to conventional intermittent hemodialysis techniques for treatment of ARF. They are particularly valuable techniques in patients in whom intermittent hemodialysis fails to control hypervolemia or uremia and for those who do not tolerate intermittent hemodialysis and in whom peritoneal dialysis is not possible. CAVH requires both arterial and venous access. The patient's own blood pressure generates an ultrafiltrate of plasma across a porous biocompatible dialysis membrane. A physiologic crystalloid solution is passed along the other side of the membrane to achieve diffusive clearance. CVVH, in contrast, requires only a double-lumen venous catheter as a blood pump generates ultrafiltration pressure across the dialysis membrane. These newer continuous techniques have not been compared to conventional intermittent hemodialysis in prospective, adequately controlled trials, and the choice of technique is currently tailored to the specific needs of the patient, the resources of the institution, and the expertise of the physician. Potential disadvantages of continuous hemodialysis techniques are the need for prolonged immobilization in bed, systemic anticoagulation, arterial cannulation (in CAVH), and prolonged exposure

of blood to synthetic, albeit relatively biocompatible, dialysis membranes.

OUTCOME AND LONG-TERM PROGNOSIS

The mortality rate among patients withARFapproximates 50% and has changed little over the past 30 years. It should be stressed, however, that patients usually die from sequelae of the primary illness that induced ARF and not from ARF itself. Indeed, the kidney is one of the few organs whose function can be replaced artificially (i.e., by dialysis) for protracted periods of time. In agreement with this interpretation, mortality rates vary greatly depending on the cause of ARF: ~15% in obstetric patients, ~30% in toxin-related ARF, and ~60% following trauma or major surgery. Oliguria (<400 mL/d) at time of presentation and a rise in serum creatinine of >265 umol/L (>3 mg/dL) are associated with a poor prognosis and probably reflect the severity of renal injury and of the primary illness. Mortality rates are higher in older debilitated patients and in those with multiple organ failure. Most patients who survive an episode of ARF recover sufficient renal function to live normal lives. However, 50% have subclinical impairment of renal function or residual scarring on renal biopsy. Approximately 5% of patients never recover function and require long-term renal replacement with dialysis or transplantation. An additional 5% suffer progressive decline inGFR, following an initial recovery phase, probably due to hemodynamic stress and sclerosis of remnant glomeruli (Chap. 273).

(Bibliography omitted in Palm version)

Back to Table of Contents

270. CHRONIC RENAL FAILURE - *Karl Skorecki*, *Jacob Green*, *Barry M. Brenner*

MECHANISMS OF CHRONIC RENAL FAILURE

DEFINITIONS

Chronic renal disease (CRD) is a pathophysiologic process with multiple etiologies, resulting in the inexorable attrition of nephron number and function, and frequently leading to *end-stage renal disease (ESRD)*. In turn, ESRD represents a clinical state or condition in which there has been an irreversible loss of endogenous renal function, of a degree sufficient to render the patient permanently dependent upon renal replacement therapy (dialysis or transplantation) in order to avoid life-threatening *uremia*. Uremia is the clinical and laboratory syndrome, reflecting dysfunction of all organ systems as a result of untreated or undertreated acute or chronic renal failure. Given the capacity of the kidneys to regain function following acute injury (Chap. 269), the vast majority (>90%) of patients with ESRD have reached this state as a result of CRD.

PATHOPHYSIOLOGY OF CRD

The pathophysiology ofCRDinvolves initiating mechanisms specific to the underlying etiology as well as a set of progressive mechanisms that are a common consequence following long-term reduction of renal mass, irrespective of etiology. Such reduction of renal mass causes structural and functional hypertrophy of surviving nephrons. This compensatory hypertrophy is mediated by vasoactive molecules, cytokines, and growth factors and is due initially to adaptive hyperfiltration, in turn mediated by increases in glomerular capillary pressure and flow. Eventually, these short-term adaptations prove maladaptive, in that they predispose to sclerosis of the remaining viable nephron population. This final common pathway for inexorable attrition of residual nephron function may persist even after the initiating or underlying disease process has become inactive. Increased intrarenal activity of the renin-angiotensin axis appears to contribute both to the initial adaptive hyperfiltration and to the subsequent maladaptive hypertrophy and sclerosis. These maladaptive long-term actions of renin-angiotensin axis activation are mediated in part through downstream growth factors such as transforming growth factor b. Interindividual variability in the risk and rate of CRD progression can be explained in part by variations in the genes encoding components of these and other pathways involved in glomerular and tubulointerstitial fibrosis and sclerosis (see "Genetic Considerations in the Progression of CRD," below).

The earliest stage common to all forms of CRDis a loss of renal reserve. When kidney function is entirely normal, glomerular filtration rate (GFR) can be augmented by 20 to 30% in response to the stimulus of a protein challenge. During the earliest stage of loss of renal reserve, basal GFR may be normal or even elevated (hyperfiltration), but the expected further rise in response to a protein challenge is attenuated. This early stage is particularly well documented in diabetic nephropathy. At this stage, the only clue may be at the level of laboratory measurements, which estimate GFR. The most commonly utilized laboratory measurements are the serum urea and creatinine concentrations. By the time serum urea and creatinine concentrations are even mildly elevated, substantial chronic nephron injury has already occurred.

AsGFRdeclines to levels as low as 30% of normal, patients may remain asymptomatic with only biochemical evidence of the decline in GFR, i.e., rise in serum concentrations of urea and creatinine. However, careful scrutiny usually reveals early additional clinical and laboratory manifestations of renal insufficiency. These may include nocturia, mild anemia and loss of energy, decreasing appetite and early disturbances in nutritional status, and abnormalities in calcium and phosphorus metabolism (*moderate renal insufficiency*). As GFR falls to below 30% of normal, an increasing number and severity of uremic clinical manifestations and biochemical abnormalities supervene (*severe renal insufficiency*). At the stages of mild and moderate renal insufficiency, intercurrent clinical stress may compromise renal function still further, inducing signs and symptoms of overt uremia. Such intercurrent clinical conditions to which patients withCRD may be particularly susceptible include infection (urinary, respiratory, or gastrointestinal), poorly controlled hypertension, hyper- or hypovolemia, and drug or radiocontrast nephrotoxicity, among others. When GFR falls below 5 to 10% of normal(ESRD), continued survival without renal replacement therapy becomes impossible.

ETIOLOGY

There has been a dramatic increase in the incidence ofESRD as well as a shift in the relative incidence of etiologies ofCRDduring the past two decades. Whereas glomerulonephritis was the leading cause of CRD in the past, diabetic and hypertensive nephropathy are now much more frequent underlying etiologies (Table 270-1). This may be a consequence of more effective prevention and treatment of glomerulonephritis or of diminished mortality from other causes among individuals with diabetes and hypertension. Greater overall longevity and diminished premature cardiovascular mortality have also increased the mean age of patients presenting with ESRD. Hypertension is a particularly common cause of CRD in the elderly, in whom chronic renal ischemia due to renovascular disease may be an underrecognized additional contribution to the pathophysiologic process. Many patients present at an advanced stage of CRD, precluding definitive determination of etiology.

GENETIC CONSIDERATIONS

Progression of CRD Disorders with clear-cut monogenic inheritance comprise a small but important component among the etiologies of CRD. Among these, autosomal dominant polycystic kidney disease is the most common on a world-wide basis (Chap. 276). Alport's hereditary nephritis (Chap. 275) is a less common cause of both benign hematuria without progression to CRD and more severe nephron injury with progression toESRD, and it usually displays an X-linked pattern of inheritance. In contrast, the two most common etiologies of CRD (Table 270-1), namely diabetes mellitus (both types 1 and 2) and essential hypertension, display complex polygenic patterns of inheritance. Both candidate locus and genome-wide strategies have been used to pinpoint genes that contribute to the risks for development of these disorders. Recent evidence also suggests that reflux nephropathy may have a heritable basis, again involving the contribution of several genetic loci.

Striking interindividual variability in the rate of progression toESRD is a characteristic feature among patients with either inherited or acquired causes of CRD, irrespective of underlying etiology. This interindividual variability has an important heritable component,

clarification of which may help guide therapeutic approaches. A number of genetic loci that contribute to the progression of CRD have been identified. Most extensively studied has been an insertion/deletion polymorphism of the angiotensin-converting enzyme (ACE) gene, previously shown to contribute to cardiovascular disease risk. Studies in a wide variety of disorders, including diabetic nephropathy, glomerulonephritis, polycystic kidney disease, and CRD caused by urologic abnormalities, have revealed an important contribution of this locus to progressive deterioration of renal function. The two different alleles defined by this polymorphism of the ACE gene are associated with corresponding differences in the endogenous activity of the encoded enzyme. The homozygous deletion (D/D) variant is associated with the highest expression of endogenous ACE activity and a greater risk of CRD progression. This finding has important therapeutic implications and leads to the prediction that ACE inhibitor therapy might be most effective in patients who are homozygous for the "at-risk" allele. Similar conclusions have been reached with respect to genes encoding other components of the renin-angiotensin axis, including the angiotensinogen gene, and the angiotensin receptor. These findings are consistent with the important role of intraglomerular hemodynamic perturbations in progressive renal injury.

PATHOPHYSIOLOGY AND BIOCHEMISTRY OF UREMIA

The uremic syndrome results from functional derangements of many organ systems, although the prominence of specific symptoms varies among patients. *Azotemia* refers to the retention of nitrogenous waste products as renal insufficiency develops. *Uremia* refers to the more advanced stages of progressive renal insufficiency when the complex, multiorgan system derangements become clinically manifest. The term uremia was adopted originally because of the presumption that all of the abnormalities result from retention in the blood of end products of metabolism normally excreted in the urine. The most likely candidates as toxins in uremia are the by-products of protein and amino acid metabolism. Unlike fats and carbohydrates, which are eventually metabolized to carbon dioxide and water -- substances readily excreted even in uremic subjects via lungs and skin -- the products of protein and amino acid metabolism depend primarily on the kidneys for excretion. Although a number of such products have been identified (Table 270-2), the clinical symptoms of uremia correlate poorly with the blood levels of these products. This is because uremia involves more than renal excretory failure alone. A host of metabolic and endocrine functions normally subserved by the kidney are also impaired, resulting in anemia; malnutrition; impaired metabolism of carbohydrates, fats, and proteins; defective utilization of energy; and metabolic bone disease. Thus, the pathophysiology of the uremic syndrome can be divided into those sets of abnormalities consequent to the accumulation of products of protein metabolism on the one hand, and on the other hand, abnormalities consequent to the loss of other renal functions, such as fluid and electrolyte homeostasis and synthesis of certain hormones [e.g., erythropoietin (EPO), 1.25-dihydroxycholecalciferol].

Although not the major cause of overt uremic toxicity, urea may contribute to some of the clinical abnormalities, including anorexia, malaise, vomiting, and headache. Elevated levels of plasma *guanidinosuccinic acid*, by interfering with activation of platelet factor III by ADP, contribute to the impaired platelet function inCRD. *Creatinine* may cause adverse effects following conversion to metabolites such as sarcosine and methylguanidine. Nitrogenous compounds with a molecular mass of 500 to 12,000 Da

(so-called middle molecules) are also retained in CRD and similarly are believed to contribute to morbidity and mortality in uremic subjects. Decreased renal excretion is not the only reason why such middle-sized molecules, along with various cytokines and growth factors, accumulate in uremic plasma. The kidney normally catabolizes a number of circulating plasma proteins and polypeptides; with reduced renal mass, this capacity is impaired. Furthermore, plasma levels of many polypeptide hormones, including parathyroid hormone (PTH) insulin, glucagon, luteinizing hormone, and prolactin, rise with renal failure, not only because of impaired renal catabolism but also because of enhanced glandular secretion. Of these, excessive PTH has been suggested to be an important uremic "toxin" because of its adverse effect of elevating cellular cytosolic Ca^{2+}levels in several tissues and organs.

CLINICAL AND LABORATORY MANIFESTATIONS OF CHRONIC RENAL FAILURE AND UREMIA

Uremia leads to disturbances in the function of every organ system. Chronic dialysis (Chap. 271) reduces the incidence and severity of these disturbances, so that, where modern medicine is practiced, the overt and florid manifestations of uremia have largely disappeared. Unfortunately, as indicated in Table 270-3, even optimal dialysis therapy is not a panacea, because some disturbances resulting from impaired renal function fail to respond fully, while others continue to progress.

FLUID, ELECTROLYTE, AND ACID-BASE DISORDERS (See also Chaps. 49 and50)

Sodium and Water Homeostasis When theGFR is normal, >24,000 mmol of Na^+ are filtered per day. An overwhelming fraction of this Na^+ load is reabsorbed by the tubules, leaving only a small fraction (usually<1%) to be excreted. Thus, even when the GFR falls markedly to levels as low as 10% of normal, the filtered load of Na^+still far exceeds daily requirements for urinary Na^+excretion. Therefore, any abnormalities in overall Na^+balance will reflect the relationship between the filtered load and fractional reabsorption (glomerulotubular balance). Progressive nephron injury can be associated with a tendency to Na^+retention, Na^+wasting, or maintenance of Na^+balance, depending in part on the underlying etiology (glomerular vs. tubulointerstitial disease), ongoing diuretic treatment, and comorbid conditions that affect Na^+balance, such as cardiac failure or cirrhosis. Osmotic regulation of vasopressin release and of thirst are also preserved. Even when tubule response to vasopressin is diminished, normal thirst mechanisms and access to H_2O generally prevent hypernatremia. However, a compromised capacity to excrete a maximally dilute urine with progressiveCRD may lead to hyponatremia.

In most patients with stableCRD, the total body contents of Na^+ and H_2O are increased modestly, although this may not be clinically apparent. The underlying etiologic disease process may itself disrupt glomerulotubular balance and promote Na^+retention (e.g., glomerulonephitis), or excessive Na^+ingestion may lead to cumulative positive Na^+balance and attendant extracellular fluid volume (ECFV) expansion. Such ECFV expansion contributes to hypertension, which in turn accelerates further the progression of nephron injury. As long as water intake does not exceed the capacity for free water clearance, the ECFV expansion will be isotonic and the patient will remain normonatremic. On the other hand, hyponatremia will be the consequence of excessive

water ingestion. However, in view of the concomitant impairment in urinary concentrating mechanism, severe hyponatremia is not usual in predialysis patients, and water restriction is only necessary when hyponatremia is documented. In such patients, a daily intake of fluid equal to the urine volume per day plus about 500 mL usually maintains the serum Na^+ concentration at normal levels.

Weight gain usually associated with volume expansion may be offset in patients with CRD by concomitant loss of lean body mass. In the CRD patient who is not yet on dialysis but has clear evidence of ECFV expansion, administration of loop diuretics coupled with restriction of salt intake are the mainstays of therapy. It should be noted that resistance to loop diuretics in renal failure often mandates use of higher doses than those usually used when GFR is well preserved. The combination of loop diuretics with metalozone, which inhibits the Na-Cl cotransporter of the distal convoluted tubule, can sometimes overcome diuretic resistance. When the GFR falls to<5 to 10 mL/min, even high doses of combination diuretics are ineffective. ECFV expansion under these circumstances usually means that dialysis is indicated. In dialysis patients with volume expansion, management should include ultrafiltration and restriction of salt and water intake between dialysis treatments.

Patients with CRD also have impaired renal mechanisms for conserving Na^+ and H_2O (Chap. 268). When an *extrarenal* cause for fluid loss is present (e.g., vomiting, diarrhea, sweating, fever), these patients are prone to volume depletion. Depletion of ECFV may compromise residual renal function with resulting signs and symptoms of overt uremia. Because of impaired renal Na^+ and H_2O conservation mechanisms, the usual indices of prerenal azotemia (oliguria, high urine osmolality, low urinary Na^+ concentration, and low fractional excretion of Na^+) are not useful. Cautious volume repletion, usually with normal saline, returns ECFV to normal and usually restores renal function to prior levels.

Potassium Homeostasis (See also Chap. 49) When GFR is normal, the approximate daily filtered load of K^+ is 700 mmol. The majority of this filtered load is reabsorbed in tubule segments prior to the cortical collecting tubule, and most of the K^+ excreted in the final urine reflects events governing K^+ handling at the level of the cortical collecting tubule and beyond. These factors include the flow of luminal fluid and the delivery and reabsorption of Na^+, which generates the lumen-negative electromotive force for K^+ secretion at the aldosterone-responsive distal nephron sites. In CRD, these factors may be well preserved, such that a decline in GFR is not necessarily accompanied by a concomitant and proportionate decline in urinary K^+ excretion. In addition, K^+ excretion in the gastrointestinal tract is augmented in patients with CRD. However, hyperkalemia may be precipitated in a number of clinical situations, including augmented dietary intake, protein catabolism, hemolysis, hemorrhage, transfusion of stored red blood cells, metabolic acidosis, and following the exposure to a variety of medications that inhibit K^+ entry into cells or K^+ secretion in the distal nephron. Most commonly encountered medications in this regard are beta blockers, ACE inhibitors, K^+-sparing diuretics (amiloride, triamterene, spironolactone), and nonsteroidal anti-inflammatory drugs (NSAIDs). In addition, certain etiologies of CRD may be associated with earlier and more severe disruption of K^+ secretory mechanisms in the distal nephron, relative to the reduction in GFR. Most important are conditions associated with hyporeninemic hypoaldosteronism (e.g., diabetic nephropathy and certain forms of distal renal tubular acidosis; Chap. 49).

Most commonly, clinically significant hyperkalemia does not occur until the GFR falls to below 10 mL/min or unless there is exposure to a K^+ load, either endogenous (e.g., hemolysis, trauma, infection) or exogenous (e.g., administration of stored blood, K^+-containing medications, K^+-containing dietary salt substitute). In kidney transplant recipients, cyclosporine is another common cause of increased plasma K^+ concentration. Hyperkalemia in CRD patients may also be induced by abrupt falls in plasma pH, since acidosis is associated with efflux of K^+ from the intracellular to the extracellular fluid compartment.

Although total-body K^+ is frequently reduced in CRD, *hypokalemia* is uncommon. The occurrence of hypokalemia usually reflects markedly reduced dietary K^+ intake, in association with excessive diuretic therapy or gastrointestinal losses. Hypokalemia occurs as a result of primary renal K^+ wasting in association with other solute transport abnormalities, as in Fanconi's syndrome, renal tubular acidosis, or other forms of hereditary or acquired tubulointerstitial diseases. However, even under these circumstances, as GFR declines, the tendency to hypokalemia diminishes and hyperkalemia may supervene. Accordingly, K^+ supplementation and K^+-sparing diuretics should be used with caution as GFR declines.

Metabolic Acidosis (See also Chap. 50) In adults, the metabolism of dietary protein generates approximately 1 mmol/kg per day of H^+. The H^+ must be excreted, primarily by renal mechanisms, if neutral acid-base balance is to be maintained. Acidosis is a common disturbance during the advanced stages of CRD. Although in a majority of patients with CRD the urine can be acidified normally, these patients have a reduced ability to produce ammonia. In part this is a consequence of limited ATP utilization, resulting from diminished Na^+ reabsorption in the proximal tubule. As a result, the use of glutamine as an energy source is limited, which in turn limits proximal tubule ammonia production. Hyperkalemia, when present, further depresses urinary ammonium excretion. The combination of hyperkalemia and hyperchloremic metabolic acidosis (known as type IV renal tubular acidosis or hyporeninemic hypoaldosteronism) is most characteristically seen in patients with diabetes or in those with predominantly tubulointerstitial disease. Treatment of the hyperkalemia frequently improves the acidosis as well.

With advancing renal failure, total urinary net daily acid excretion is usually limited to 30 to 40 mmol; thus, throughout the remainder of their course of CRD, many patients may be in a positive H^+ balance of 20 to 40 mmol/d. The retained H^+ is buffered by bone salts. In the early stages, the accompanying organic anions are excreted in the urine, and the metabolic acidosis is of the non-anion gap variety. However, with advanced renal failure, a fairly large "anion gap" may develop (to approximately 20 mmol/L) with a reciprocal fall in plasma HCO_3^- concentration. In most patients, the metabolic acidosis is mild and the pH is rarely less than 7.35. The metabolic acidosis can usually be corrected by treating the patient with 20 to 30 mmol of sodium bicarbonate or sodium citrate daily. However, the concomitant Na^+ load mandates careful attention to volume status and the potential need for diuretic agents. Also, citrate enhances aluminum absorption in the large bowel, and citrate-containing antacids should be avoided if aluminum-containing drugs are also administered. As with other abnormalities in CRD, severe symptomatic manifestations of acid-base imbalance occur when the patient is

challenged with an excessive endogenous or exogenous acid load or loses excessive alkali (e.g., with diarrhea).

BONE, PHOSPHATE, AND CALCIUM ABNORMALITIES (Fig. 270-1) (See also Chap. 340)

Although clinical symptoms of bone disease are present before dialysis in fewer than 10% of patients with ESRD, radiologic and histologic abnormalities are observed in about 35 and 90%, respectively. Two principal types of bone disorders are observed in patients with ESRD: a high-turnover osteodystrophy, known as *osteitis fibrosa cystica*, and a low-turnover state characterized initially by *osteomalacia* and subsequently by *adynamic bone disease*. In osteitis fibrosa, the number and size of the osteoclasts are increased, as are the number and depth of the osteoclastic resorption lacunae. Collagen deposition is less ordered, and the rate of bone turnover is markedly increased. In osteomalacia, the rate of mineralization is slower than that of collagen synthesis, resulting in excessive accumulation of unmineralized osteoid and widened osteoid seams. In adynamic uremic osteodystrophy, a parallel marked reduction in the rate of mineralization and collagen synthesis results in osteoid seams of normal width. While these disorders are often discussed as if they were distinct, they commonly overlap in a given patient with ESRD.

High-Turnover Uremic Osteodystrophy This condition is associated with elevated PTH levels. The hyperparathyroid state is attributable both to hyperplastic growth of the parathyroid glands and to augmented release of hormone from each individual parathyroid cell. The main factors responsible for deranged PTH synthesis in CRD are related to altered metabolism of phosphate, calcitrol [1,25(OH)$_2$D$_3$], and Ca^{2+}.

Phosphate (PO_4^{3-}) Hyperphosphatemia is a feature of advanced renal failure. The serum phosphate concentration rises in patients with a GFR < 20 mL/min, but retention of PO_4^{3-} can be documented in balance studies with even less severe declines in GFR. The retained PO_4^{3-} is a major cause of the development of secondary hyperparathyroidism in CRD. PO_4^{3-} exerts indirect effects on PTH secretion by decreasing renal production of calcitriol (see below) and by lowering plasma ionized Ca^{2+}. Recent studies also suggest a direct stimulatory role of PO_4^{3-} at the level of the parathyroid gland, in the absence of changes in serum Ca^{2+} or calcitriol levels. Dietary restriction of PO_4^{3-} as well as gastrointestinal PO_4^{3-} binders may prevent hyperphosphatemia, thereby mitigating the rise in PTH levels.

Calcitriol Under normal conditions, calcitriol exerts negative feedback control on the parathyroid gland through both direct (i.e., diminished transcription of pre-pro PTH mRNA) and indirect mechanisms. The latter act through stimulation of intestinal absorption of Ca^{2+} and the skeletal mobilization of Ca^{2+}, thereby increasing plasma Ca^{2+} and inhibiting PTH secretion. Therefore, reduced synthesis of 1,25(OH)$_2$D$_3$ during CRD plays a key role in the pathogenesis of hyperparathyroidism, both directly and through hypocalcemia. The abnormal vitamin D metabolism may be related to the renal disease itself (since the active vitamin D metabolite is normally produced in the proximal tubule) and to the hyperphosphatemia, which has a suppressive effect on the renal 1a-hydroxylase enzyme. Furthermore, a decrease in the number of calcitriol receptors in the parathyroid tissue of uremic patients has been

reported by several groups. Recent studies have demonstrated a marked decrease in vitamin D-receptor expression in areas of nodular transformation within hyperplastic parathyroid tissue but revealed no such receptor downregulation in diffuse hyperplastic parathyroid tissue. Since vitamin D also has an antiproliferative effect on parathyroid cells, this phenomenon may provide an explanation for the marked PTH secretion as well as the abnormal glandular growth pattern characteristic of nodular hyperparathyroidism.

Calcium The total plasma Ca^{2+}concentration in patients withCRD is often significantly lower than normal. Patients with CRD tolerate the hypocalcemia quite well; rarely is a patient symptomatic from the decreased Ca^{2+}concentration. This may partly be due to the frequent concomitant acidosis, which offsets some of the neuromuscular effects of hypocalcemia. The hypocalcemia in CRD results from decreased intestinal absorption of Ca^{2+} due to vitamin D deficiency (see above). Also, with the increasing serum PO_4^{3-}level, Ca^{2+}phosphate is deposited in soft tissues and serum Ca^{2+}concentration (both total and ionized) declines. In addition, patients with CRD are resistant to the action ofPTH. Hypocalcemia is a potent stimulus to PTH secretion and leads to hyperplasia of the parathyroid gland. Ca^{2+}binds to a specific Ca^{2+}-sensing receptor protein located in the cell membrane. The Ca^{2+}-sensing receptor is linked to several cytoplasmic messenger systems by one or more GTP-binding proteins. These signaling pathways are responsible for either enhanced or suppressed release of PTH during acute hypo- and hypercalcemia, respectively. Several studies have demonstrated the mRNA and protein expression of the Ca^{2+}-sensing receptor to be reduced in primary (adenomas) and secondary hyperparathyroidism (hyperplasia) compared to the expression in normal parathyroid tissue. In secondary hyperparathyroidism, expression of the Ca^{2+}-sensing receptor is often depressed in nodular areas compared with adjacent nonnodular hyperplasia. Thus, decreased Ca^{2+}receptor expression in hyperparathyroidism is compatible with a less efficient control of PTH synthesis and release, in response to varying plasma Ca^{2+}concentration.

In addition to excessive release of PTHfrom individual parathyroid cells, the size of the glands also increases as renal failure progresses. This abnormal growth of the parathyroid glands may assume one of the following patterns: (1) diffuse hyperplasia (polyclonal growth), (2) nodular growth (monoclonal growth) within diffuse hyperplastic tissue, or (3) diffuse monoclonal hyperplasia ("adenoma," or tertiary autonomous hyperparathyroidism). Patients with monoclonal ("autonomous") hyperplasia are especially prone to develop hypercalcemia following successful kidney transplantation, often necessitating parathyroidectomy.

During the initial phase ofCRD, the elevatedPTHlevels may normalize serum levels of Ca^{2+}, PO_4^{3-}, and vitamin D. Therefore hypocalcemia, hyperphosphatemia, and reduced $1,25(OH)_2D_3$are observed only as CRD progresses. However, even at the earliest stages of CRD, the elevated PTH levels adversely affect bone metabolism, causing increased osteoclastic and osteoblastic activity (high-turnover bone disease). Additional detrimental factors include the chronic uremic acidosis, which inhibits osteoblastic bone formation and stimulates osteoclastic bone resorption.

Low-Turnover Uremic Osteodystrophy Originally thought to result solely from vitamin D deficiency, *osteomalacia* (Chap. 342) has now been more closely associated with

aluminum toxicity. Aluminum was first identified as a presumed cause of dialysis dementia in dialysis patients, and shortly thereafter aluminum deposition in bone was shown to be associated with osteomalacia. The sources of aluminum were phosphate binders and the water used in preparing dialysate. Aluminum is no longer present as a contaminant in dialysate, but it still is widely utilized as a phosphate binder in some settings. Approximately one-third of dialysis andCRDpatients ingest at least some aluminum. Aluminum deposition adversely affects mineral deposition at the mineralization front.

Aplastic renal osteodystrophy occurs in many patients who have no evidence of excess aluminum accumulation. These patients have relatively low levels ofPTH. The disorder is associated with the use of supraphysiologic Ca^{2+}concentrations in peritoneal dialysate and the excessive use of oral Ca^{2+} and vitamin D preparations in both hemodialysis and peritoneal dialysis patients. These sources of exogenous Ca^{2+}might lower serum PTH to levels that are inadequate for maintaining normal bone turnover.

Yet another type of skeletal lesion that occurs inESRDpatients after many years of dialysis therapy results from *amyloid deposition* related to the accumulation ofb_2-microglobulin. This syndrome presents as carpal tunnel syndrome, tenosynovitis of the hands, shoulder arthropathy, bone cysts, cervical spondyloarthropathy, and cervical pseudotumors. It is characterized on x-ray films by cysts in the carpal bones and femoral neck. Amyloid tumoral masses may be best appreciated by ultrasound examination or computed tomography.

With high-turnover osteitis fibrosa cystica, vitamin D-deficient and aluminum-induced osteomalacia, and dialysis-related amyloidosis,ESRDpatients are prone to spontaneous fractures, which are slow to heal. The ribs are most commonly involved in the case of osteitis fibrosa cystica. The femoral neck is a frequent site of aluminum-induced osteomalacia and dialysis-related amyloidosis and is also prone to pathologic fractures. Bone pain, even in the absence of fractures, is common. In osteitis fibrosa cystica, a proximal myopathy often coexists, giving rise to gait abnormalities and to impairment of ambulation. Similarly, a myopathy may also accompany amyloid arthropathy. InCRD, there is a tendency to extraosseous or metastatic calcification when the calcium-phosphate product is very high (>70 when expressed as mg/dL). Medium-sized blood vessels; subcutaneous, articular, and periarticular tissues; myocardium; eyes; and lungs are common sites of metastatic calcification. *Calciphylaxis* refers to devastating necrotic extremity soft tissue lesions related to vascular occlusion and metastatic calcification.

The Effect of Uremic Acidosis on Bone Disease As previously noted, in patients withCRD, a decrease in acid excretion leads to unremitting positive H^+balance. If extracellular fluid HCO_3^-were the only H^+buffer available, it would become progressively depleted and the concentration of serum HCO_3^-, and thus pH, would fall to levels incompatible with life. However, during CRD, extracellular fluid $HCO_3^-$and pH remain stable, although reduced, for long periods; thus, either non-HCO_3^-buffers must neutralize the retained hydrogen ions or acid production must decrease. Acid production does not appear to diminish in patients with renal failure; yet such patients excrete only approximately two-thirds of their daily hydrogen ion production. Thus substantial buffering of the retained hydrogen ions almost certainly occurs. Because of its mass and

potential buffering capacity, bone is a likely site for the chronic hydrogen ion buffering.

TREATMENT

Secondary hyperparathyroidism and *osteitis fibrosa* are best prevented and treated by reducing serum PO_4^{3-} concentration through the use of a PO_4^{3-}-restricted diet as well as oral PO_4^{3-}-binding agents. Calcium carbonate and calcium acetate are the preferred PO_4^{3-}-binding agents, but in some rare circumstances a combination of short-term aluminum hydroxide and calcium carbonate is necessary. In such cases, aluminum levels should be monitored, and citrate antacids, which enhance aluminum absorption, should be avoided. Daily oral calcitriol, or intermittent oral or intravenous pulses, appear to exert a direct suppressive effect on PTH secretion, in addition to the indirect effect mediated through raising Ca^{2+} levels. Intravenous pulses are especially convenient for patients on hemodialysis. The use of calcitriol and Ca^{2+} preparations in the predialysis population must take into account potential effects of increased phosphate and Ca^{2+} on the rate of progression of CRD. In the dialysis population, dialysate Ca^{2+}, calcium carbonate, calcium acetate, aluminum hydroxide, and calcitriol must be properly balanced to maintain the serum PO_4^{3-} concentration at approximately 1.4 mmol/L (4.5 mg/dL) and the serum Ca^{2+} at approximately 2.5 mmol/L (10 mg/dL) in an attempt to suppress parathyroid hyperplasia, thus avoiding or reversing osteitis fibrosa cystica, osteomalacia, and myopathy. It is particularly important to maintain the Ca^{2+}-PO_4^{3-} product in the normal range to avoid metastatic calcification. Several analogues of calcitriol are now being evaluated. Such analogues would be beneficial if they had the same effect on PTH mRNA as calcitriol and increased the margin of safety and efficacy by having less hyperphosphatemic and hypercalcemic effects. In this manner, it might be possible to have a greater suppressive effect on PTH transcription because a higher dose could be used safely.

Adynamic bone disease is often a consequence of overzealous treatment of secondary hyperparathyroidism. Therefore, suppression of PTH levels to less than 120 pg/mL in uremic patients may not be desirable. The incidence of aluminum-induced osteomalacia has been greatly reduced with the recognition of aluminum as the principal culprit. Therapy for this disorder is continued avoidance of aluminum, with possible use of a chelating agent such as desferoxamine along with high-flux dialysis. Management of metabolic acidosis should aim to maintain a nearly normal level of plasma HCO_3^-, with the administration of calcium acetate or calcium carbonate in the first instance, and with the addition of $NaHCO_3$ if necessary. Excessive administration of alkali should be avoided to minimize risk of urinary precipitation of calcium phosphate.

At present, there is no good therapy for *dialysis-related amyloidosis*. Local physical therapy, glucocorticoid injections, and NSAIDs constitute current options.

Other Solutes *Uric acid* retention is a common feature of CRD but rarely leads to symptomatic gout. Treatment of hyperuricemia is not necessary unless recurrent gout becomes a problem. When recurrent symptomatic gout occurs, a reduced dose of allopurinol (100 to 200 mg/d) is usually sufficient to inhibit uric acid synthesis. Hypophosphatemia is rare and, when it occurs, is usually a consequence of overzealous oral administration of phosphate-binding gels. Because serum magnesium levels tend to rise in CRD, magnesium-containing antacids and cathartics should be

avoided.

CARDIOVASCULAR AND PULMONARY ABNORMALITIES

Congestive Heart Failure (See also Chap. 232) Salt and water retention in uremia often result in congestive heart failure and/or pulmonary edema. A unique form of pulmonary congestion and edema may occur even in the absence of volume overload and is associated with normal or mildly elevated intracardiac and pulmonary capillary wedge pressures. This entity, characterized radiologically by peripheral vascular congestion giving rise to a "butterfly wing" distribution, is due to increased permeability of alveolar capillary membranes. This "low-pressure" pulmonary edema as well as cardiopulmonary abnormalities associated with circulatory overload usually respond promptly to vigorous dialysis.

Hypertension and Left Ventricular Hypertrophy (See also Chap. 246) Hypertension is the most common complication ofCRD andESRD. When it is not found, the patient may have a salt-wasting form of renal disease (e.g., medullary cystic disease, chronic tubulointerstitial disease, or papillary necrosis), may be receiving antihypertensive therapy, or be volume-depleted, the last condition usually due to excessive gastrointestinal fluid losses or overzealous diuretic therapy. Since volume overload is the major cause of hypertension in uremia, the normotensive state can often be restored by appropriate use of diuretics in the predialysis patient or with aggressive ultrafiltration in dialysis patients. Nevertheless, because of hyperreninemia, some patients remain hypertensive despite rigorous salt and water restriction and ultrafiltration. Rarely, patients develop accelerated or malignant hypertension. Intravenous nitroprusside, labetolol, or more recently approved agents such as fenoldopam or urapidil, together with control ofECFV, generally controls such hypertension. Subsequently, such patients usually require more than one oral antihypertensive drug. Enalaprilat or otherACEinhibitors may also be considered, but in the face of bilateral renovascular disease they have the potential to further reduceGFRabruptly. Administration of erythropoietin (EPO) (p. 1557) may raise blood pressure and increase the requirement for antihypertensive drugs. A high percentage of patients with CRD present with left ventricular hypertrophy or dilated cardiomyopathy. These are among the most ominous risk factors for excess cardiovascular morbidity and mortality in patients with CRD and ESRD and are thought to be related primarily to prolonged hypertension and ECFV overload. In addition, anemia and the surgical placement of an arteriovenous anastomosis for future or ongoing dialysis access may generate a high cardiac output state, which also intensifies the burden placed on the left ventricle.

TREATMENT

Management of hypertension inCRD can be considered in terms of two overall goals: to slow the progression of CRD itself and to prevent the extrarenal complications of hypertension, such as cardiovascular disease and stroke. In all patients with CRD, blood pressure should be controlled to at least the level established in the guidelines of the Joint National Commission on Hypertension Detection Education and Follow-up Program (130/80-85 mmHg). In the elderly, levels of 140 mmHg may be a more realistic target. In predialysis patients with proteinuria> 1 g per 24 hr, blood pressure should be further reduced to a mean arterial pressure of 92 mmHg (equivalent to 125/75 mmHg),

where possible. Volume control is the mainstay of therapy, with addition of antihypertensive agents as needed when hypertension persists despite achievement of a normovolemic state. When salt retention and hypervolemia contribute to hypertension, salt restriction and diuretics are indicated as initial therapy. The choice of additional agents to slow the progression of CRD based upon reduction of intraglomerular hypertension and proteinuria is considered below. In ESRDpatients, considerations related to slowing of nephron injury are less important, and the main goal is to prevent cardiac hypertrophy and stroke. The choice of antihypertensive agents may come from all the major classes, with careful consideration of comorbid conditions. However, powerful direct-acting vasodilators, such as hydralazine or minoxidil, may perpetuate the tendency to cardiac hypertrophy, despite the lowering of blood pressure. Therefore, prolonged use of such agents should be reserved for those very rare patients in whom severe refractory hypertension persists, despite adequate volume reduction and compliance with all other classes of antihypertensives.

Atherosclerotic Coronary and Peripheral Vascular Disease Hypertension, hyperhomocysteinemia, and lipid abnormalities promote atherosclerosis but are potentially treatable complications ofCRD. Ongoing or prior nephrotic syndrome is also associated with hyperlipidemia and hypercoagulability, which increase the risk of occlusive vascular disease. Since diabetes mellitus and hypertension are themselves the two most frequent etiologies of CRD, it is not surprising that cardiovascular disease is the most frequent cause of death inESRDpatients. Therefore, accepted life-style changes and therapeutic measures for cardiac risk reduction (Chap. 242) are especially important in this group of patients. The approach to managing hypertension has been outlined above. Hyperhomocysteinemia may respond to vitamin therapy, which includes folate supplementation to between 1 and 5 mg/d. Hyperlipidemia in patients with CRD and ESRD should be managed aggressively according to the guidelines of the National Cholesterol Education Program (Chaps. 242 and344). If dietary measures are inadequate, the preferred lipid-lowering medications are gemfibrozil and an HMG-CoA reductase inhibitor. However, these two classes of agents ordinarily should not be combined because of an increased risk of myositis and rhabomyolysis in CRD and ESRD patients.

Abnormalities in Ca^{2+} and PO_4^{3-}metabolism (see above) may lead to metastatic vascular calcification and markedly increase the propensity to coronary, cerebral, and peripheral occlusive vascular disease. By careful attention to the guidelines noted above for the management of divalent ion metabolism and bone disease, avoidance of an elevated Ca^{2+}-PO_4^{3-}product may mitigate this effect.

Pericarditis (See alsoChap. 239) With the advent of early initiation of renal replacement therapy, pericarditis is now observed more often in underdialyzed patients than in patients withCRD in whom dialysis has not yet been initiated. Pericardial pain with respiratory accentuation, accompanied by a friction rub, are the hallmarks of uremic pericarditis. The finding of a multicomponent friction rub strongly supports the diagnosis. Furthermore, the usual occurrence of multiple cardiac murmurs, S_3 and S_4 heart sounds, and transmitted bruits from arteriovenous access devices may render precordial auscultation more challenging in this group of patients. Classic electrocardiographic abnormalities include PR-interval shortening and diffuse ST-segment elevation. Pericarditis may be accompanied by the accumulation of

pericardial fluid, readily detected by echocardiography, sometimes leading to cardiac tamponade. Pericardial fluid in uremic pericarditis is more often hemorrhagic than in viral pericarditis.

TREATMENT

Uremic pericarditis is an absolute indication for initiation of dialysis or for intensification of the dialysis prescription in those already on dialysis. Because of the propensity to hemorrhagic pericardial fluid, heparin-free dialysis is indicated. Pericardiectomy should be considered only if more conservative measures fail. Nonuremic causes of pericarditis and pericardial effusion include viral, malignant, and tuberculous pericarditis and pericarditis associated with myocardial infarction; these are also more frequent in patients withESRD and should be managed according to the dictates of the underlying disease process.

HEMATOLOGIC ABNORMALITIES

Anemia ofCRD(See also Chap. 105) A normocytic, normochromic anemia is present in the majority of patients withCRD. It is usually observed when theGFRfalls below 30 mL/min. When untreated, the anemia of CRD is associated with a number of physiologic abnormalities, including decreased tissue oxygen delivery and utilization, increased cardiac output, cardiac enlargement, ventricular hypertrophy, angina, congestive heart failure, decreased cognition and mental acuity, altered menstrual cycles, and impaired immune responsiveness. In addition, anemia may play a role in growth retardation in children. The primary cause of anemia in patients with CRD is insufficient production ofEPO by the diseased kidneys. Additional factors include the following: iron deficiency, either related to or independent of blood loss from repeated laboratory testing, needle punctures, blood retention in the dialyzer and tubing, or gastrointestinal bleeding; severe hyperparathyroidism; acute and chronic inflammatory conditions; aluminum toxicity; folate deficiency; shortened red cell survival; hypothyroidism; and underlying hemoglobinopathies. These potential contributing factors should be considered and addressed.

Before 1989, theEPO-deficient condition characteristic ofCRDcould only be treated with blood transfusions and anabolic steroids, with limited success and substantial complications. The availability of recombinant human EPO, approved by the U.S. Food and Drug Administration in 1989, has been one of the most significant advances in the care of renal patients in the past decade. Considerable debate continues regarding the optimal target hematocrit in dialysis patients receiving EPO. Mortality and hospitalization studies support the National Kidney Foundation Dialysis Outcomes Quality Initiative target hematocrit range of 33 to 36% as providing the best associated outcomes. EPO can be administered either intravenously or subcutaneously. Most studies have shown that administering EPO by the subcutaneous route has a sparing effect, with the target hematocrit achieved at a lower EPO dose. ManagementGuidelinesfor the correction of anemia in CRD are as follows.

The iron status of the patient withCRD must be assessed, and adequate iron stores should be available before treatment withEPO is initiated. Iron supplementation is usually essential to ensure an adequate response to EPO in patients with CRD,

because the demands for iron by the erythroid marrow frequently exceed the amount of iron that is immediately available for erythropoiesis (as measured by percent transferrin saturation) as well as iron stores (as measured by serum ferritin). In most cases, intravenous iron will be required to achieve and/or maintain adequate iron. However, excessive iron therapy may be associated with a number of complications, including hemosiderosis, accelerated atherosclerosis, increased susceptibility to infection, and possibly an increased propensity to the emergence of malignancies. In addition to iron, an adequate supply of the other major substrates and cofactors for erythrocyte production must be assured, especially vitamin B_{12} and folate. Anemia resistant to recommended doses of EPO in the face of adequate availability of iron and vitamin factors often suggests inadequate dialysis; uncontrolled hyperparathyroidism; aluminum toxicity; chronic blood loss or hemolysis; and associated hemoglobinopathy, malnutrition, chronic infection, multiple myeloma, or another malignancy. Blood transfusions may contribute to suppression of erythropoiesis in CRD; because they increase the risk of hepatitis, hemosiderosis, and transplant sensitization, they should be avoided unless the anemia fails to respond to EPO and the patient is symptomatic.

Abnormal Hemostasis Abnormal hemostasis is common inCRD and is characterized by a tendency to abnormal bleeding and bruising. Bleeding from surgical wounds and spontaneous bleeding into the gastrointestinal tract, pericardial sac, or intracranial vault (in the form of subdural hematoma or intracerebral hemorrhage) are of greatest concern. Prolongation of bleeding time, decreased activity of platelet factor III, abnormal platelet aggregation and adhesiveness, and impaired prothrombin consumption contribute to the clotting defects. The abnormality in platelet factor III correlates with increased plasma levels of guanidinosuccinic acid and can be corrected by dialysis. Prolongation of the bleeding time is common even in well-dialyzed patients. Abnormal bleeding times and coagulopathy in patients with renal failure may be reversed with desmopressin, cryoprecipitate, conjugated estrogens, and blood transfusions, as well as by the use ofEPO.

Enhanced Susceptibility to Infection Changes in leukocyte formation and function in uremia lead to enhanced susceptibility to infection. Lymphocytopenia and atrophy of lymphoid structures occur, whereas neutrophil production is relatively unimpaired. Nevertheless, the function of all leukocyte cell types may be affected adversely by uremic serum. Alterations in monocyte, lymphocyte, and neutrophil function cause impairment of acute inflammatory responses, decreased delayed hypersensitivity, and altered late immune function.

There is a tendency for uremic patients to have less fever in response to infection, perhaps because of the effects of uremia on the hypothalamic temperature control center. Leukocyte function may also be impaired in patients withCRDbecause of coexisting acidosis, hyperglycemia, protein-calorie malnutrition, and serum and tissue hyperosmolarity (due to azotemia). In patients treated with hemodialysis, leukocyte function is disturbed because of the effects of the bioincompatibility of various dialysis membranes. Activation of cytokine and complement cascades likewise occurs when blood comes in contact with dialysis membranes. These substances in turn alter inflammatory and immune responses of the uremic patient. Mucosal barriers to infection may also be defective, and, in dialysis patients, vascular and peritoneal access devices are common portals of entry for pathogens, especially staphylococci. Glucocorticoids

and immunosuppressive drugs used for various renal diseases and renal transplantation further increase the risk of infection.

NEUROMUSCULAR ABNORMALITIES

Subtle disturbances of central nervous system function, including inability to concentrate, drowsiness, and insomnia, are among the early symptoms of uremia. Mild behavioral changes, loss of memory, and errors in judgment soon follow and may be associated with neuromuscular irritability, including hiccoughs, cramps, and fasciculations/twitching of muscles. Asterixis, myoclonus, and chorea are common in terminal uremia, as are stupor, seizures, and coma. Peripheral neuropathy is also common in advancedCRD. Initially, sensory nerves are involved more than motor nerves, lower extremities more than upper, and distal portions of the extremities more than proximal. The "restless legs syndrome" is characterized by ill-defined sensations of discomfort in the feet and lower legs requiring frequent leg movement. If dialysis is not instituted soon after onset of sensory abnormalities, motor involvement follows, including loss of deep tendon reflexes, weakness, peroneal nerve palsy (foot drop), and, eventually, flaccid quadriplegia. Accordingly, evidence of peripheral neuropathy is a firm indication for the initiation of dialysis or transplantation. Some of the central nervous system and neuromuscular complications of advanced uremia resolve with dialysis, although nonspecific electroencephalographic abnormalities may persist (Table 270-3). Successful transplantation may reverse residual peripheral neuropathy.

Two types of neurologic disturbances are unique to patients on chronic dialysis (Chap. 271). *Dialysis dementia* may occur in patients who have been on dialysis for many years and is characterized by speech dyspraxia, myoclonus, dementia, and eventually seizures and death. Aluminum intoxication is probably the major contributor to this syndrome, but other factors, such as viral infections, may play a role since not all patients with aluminum exposure develop the syndrome. *Dialysis disequilibrium*, which occurs during the first few dialyses in association with rapid reduction of blood urea levels, manifests clinically with nausea, vomiting, drowsiness, headache, and, rarely, seizures. The syndrome has been attributed to cerebral edema and increased intracranial pressure due to the rapid (dialysis-induced) shifts of omsolality and pH between extracellular and intracellular fluids. This complication can often be anticipated and prevented in patients who present with markedly elevated concentrations of plasma urea, by prescribing an initial dialysis regimen that produces slower solute removal.

GASTROINTESTINAL ABNORMALITIES

Anorexia, hiccoughs, nausea, and vomiting are common early manifestations of uremia. Protein restriction is useful in diminishing nausea and vomiting late in the course of renal failure. However, protein restriction should not be implemented in patients with early signs of protein-calorie malnutrition. *Uremic fetor*, a uriniferous odor to the breath, derives from the breakdown of urea to ammonia in saliva and is often associated with an unpleasant metallic taste sensation. Mucosal ulcerations leading to blood loss can occur at any level of the gastrointestinal tract in the very late stages ofCRD. Peptic ulcer disease is common in uremic patients. Whether this high incidence is related to altered gastric acidity, enhanced colonization by *Helicobacter pylori*, or hypersecretion of gastrin is unknown. Patients with CRD, particularly those with polycystic kidney disease,

have an increased incidence of diverticulosis. Pancreatitis and angiodysplasia of the large bowel with chronic bleeding have been noted more commonly in dialysis patients. Hepatitis B antigenemia was very common in the past, but it is much less so now because of the implementation of universal precautions, the use of hepatitis B vaccine, and the diminished need for blood transfusions resulting from the introduction of EPO. There is a higher incidence of hepatitis C virus infection in patients treated with chronic hemodialysis. Unlike hepatitis B, this infection is most often persistent. Although it does not seem to cause significant liver disease in most patients, it is a definite concern in patients who subsequently undergo transplantation and immunosuppression, in whom the incidence of active chronic hepatitis and cirrhosis is considerably higher than in those without hepatitis C infection.*The role of interferon and antiviral treatment in both hepatitis B and C infections is discussed in Chap. 295.

ENDOCRINE-METABOLIC DISTURBANCES

Disturbances in parathyroid function, protein-calorie and lipid metabolism, and overall nutritional abnormalities of uremia have already been considered.

Glucose metabolism is impaired, as evidenced by a slowing of the rate at which blood glucose levels decline after a glucose load. Fasting blood glucose is usually normal or only slightly elevated, and the mild glucose intolerance related to uremia per se, when present, does not require specific therapy. Because the kidney contributes significantly to insulin removal from the circulation, plasma levels of insulin are slightly to moderately elevated in most uremic subjects, both in the fasting and post-prandial states. However, the response to insulin and glucose utilization is impaired inCRD. Many renal hypoglycemic drugs require dose reduction in renal failure, and some, such as metformin, are contraindicated whenGFR has diminished by more than approximately 25 to 50%.

In women, *estrogen levels* are low, and amenorrhea and inability to carry pregnancies to term are common manifestations of uremia. WhenGFR has declined by approximately 30%, pregnancy may hasten the progression ofCRD. In women withESRD, the reappearance of menses is a sign of efficient renal replacement therapy and is a frequent occurrence after an adequate chronic dialysis regimen has been established. Successful pregnancies are rare. In men with CRD, including those receiving chronic dialysis, impotence, oligospermia, and germinal cell dysplasia are common, as are reduced plasma testosterone levels. Like growth, sexual maturation is often impaired in adolescent children with CRD, even among those treated with chronic dialysis. Many of these abnormalities improve or reverse with successful renal transplantation.

DERMATOLOGIC ABNORMALITIES

The skin may show evidence of anemia (pallor), defective hemostasis (ecchymoses and hematomas), calcium deposition and secondary hyperparathyroidism (pruritus, excoriations), dehydration (poor skin turgor, dry mucous membranes), and the general cutaneous consequences of protein-calorie malnutrition. A sallow, yellow cast may reflect the combined influences of anemia and retention of a variety of pigmented metabolites, or *urochromes*. The gray to bronze discoloration of the skin related to

transfusional hemochromatosis has now become uncommon with the availability and usage ofEPO. In advanced uremia, the concentration of urea in sweat may be so high that, after evaporation, a fine white powder can be found on the skin surface -- so-called uremic (urea) frost. Although many of these cutaneous abnormalities improve with dialysis, *uremic pruritus* often remains a problem. The first lines of management are to rule out unrelated skin disorders, to adjust the dialysis prescription so as to ensure adequacy of dialysis, and to control PO_4^{3-}concentration with avoidance of an elevated Ca^{2+}-PO_4^{3-}product. Occasionally, pruritus remains refractory to these measures and to other nonspecific systemic and topical therapies. The latter has itself been reported to improve pruritus. Skin necrosis can occur as part of the calciphylaxis syndrome, which also includes subcutaneous, vascular, joint, and visceral calcification in patients with poorly controlled calcium-phosphate product.

DIAGNOSTIC APPROACH

The most important initial step in the evaluation of a patient presenting de novo with biochemical or clinical evidence of renal failure is to distinguishCRD, which may be first coming to clinical attention, from true acute renal failure. The demonstration of evidence of chronic metabolic bone disease and anemia and the finding of bilaterally reduced kidney size by imaging studies strongly favor a long-standing process consistent with CRD. However, these findings do not rule out the superimposition of an acute and reversible exacerbating factor that has accelerated the decline inGFR (see below). Having established that the patient suffers from CRD, in the early stages it is often possible to establish the underlying etiology. However, when the CRD process is quite advanced, then definitively establishing an underlying etiology becomes less feasible in many cases and also of less therapeutic significance.

ESTABLISHING THE ETIOLOGY

Of special importance in establishing the etiology ofCRD are a history of hypertension; diabetes; systemic infectious, inflammatory, or metabolic diseases; exposure to drugs and toxins; and a family history of renal and urologic disease. Drugs of particular importance include analgesics (usage frequently underestimated or denied by the patient),NSAIDs, gold, penicillamine, antimicrobials, lithium, andACEinhibitors. In evaluating the uremic syndrome, questions about appetite, diet, nausea, vomiting, hiccoughing, shortness of breath, edema, weight change, muscle cramps, bone pain, mental acuity, and activities of daily living are especially helpful.

Physical Examination Particular attention should be paid to blood pressure, fundoscopy, precordial examination, examination of the abdomen for bruits and palpable renal masses, extremity examination for edema, and neurologic examination for the presence of asterixis, muscle weakness, and neuropathy. In addition, the evaluation of prostate size in men and potential pelvic masses in women should be undertaken by appropriate physical examination.

Laboratory Investigations These should also focus on a search for clues to an underlying disease process and its continued activity. Therefore, if the history and physical examination warrant, immunologic tests for systemic lupus erythematosus and vasculitis might be considered. Serum and urinary protein electrophoresis should be

undertaken in all patients over the age of 40 with unexplainedCRD and anemia, to rule out paraproteinemia. Other tests to determine the severity and chronicity of the disease include serial measurements of serum creatinine and blood urea nitrogen, hemoglobin, calcium, phosphate, and alkaline phosphatase to assess metabolic bone disease. Urine analysis may be helpful in assessing the presence of ongoing activity of the underlying inflammatory or proteinuric disease process, and when indicated should be supplemented by a 24-h urine collection for quantifying protein excretion. The latter is particularly helpful in guiding management strategies aimed at ameliorating the progression of CRD. The presence of *broad casts* on examination of the urinary sediment is a nonspecific finding seen with all diverse etiologies and reflects chronic tubulointerstitial scarring and tubular atrophy with widened tubule diameter, usually signifying an advanced stage of CRD.

Imaging Studies The most useful among these is renal sonography. An ultrasound examination of the kidneys verifies the presence of two symmetric kidneys, provides an estimate of kidney size, and rules out renal masses and obstructive uropathy. The documentation of symmetric small kidneys supports the diagnosis of progressiveCRD with an irreversible component of scarring. The occurrence of normal kidney size suggests the possibility of an acute rather than chronic process. However, polycystic kidney disease, amyloidosis, and diabetes may lead to CRD with normal-sized or even enlarged kidneys. Documentation of asymmetric kidney size suggests either a unilateral developmental or urologic abnormality or chronic renovascular disease. In the latter case, a vascular imaging procedure, such as duplex Doppler sonography of the renal arteries, radionuclide scintigraphy, or magnetic resonance angiography should be considered. A computed tomographic scan without contrast may be useful in assessing kidney stone activity, in the appropriate clinical context. Voiding cystourethrography to rule out reflux may be indicated in some younger patients with a history of enuresis or with a family history of reflux. However, in most cases, by the time CRD is established, reflux has resolved; even if present, its repair may not stabilize renal function. In any case, imaging studies should avoid exposure to intravenous radiocontrast dye where possible because of its nephrotoxicity.

Differentiation of CRD from Acute Renal Failure The most classic constellation of laboratory and imaging findings that distinguishes progressiveCRD from acute renal failure are bilaterally small (<8.5 cm) kidneys, anemia, hyperphosphatemia and hypocalcemia with elevatedPTHlevels, and a urinary sediment that is inactive or reveals proteinuria and broad casts. Furthermore, integration of a particular constellation of clinical, laboratory, and imaging findings based on the approach noted above strongly supports a particular presumed underlying etiologic disease process. For example, in a patient with insulin-dependent type 1 diabetes mellitus of 15 to 20 years' duration, diabetic retinopathy, and nephrotic-range albuminuria without hematuria, the diagnosis of diabetic nephropathy is likely. The diagnosis of chronic hypertensive nephrosclerosis (Chap. 278) requires a history of long-standing hypertension, in the absence of evidence for another renal disease process, and hence is usually a diagnosis of exclusion. Usually, proteinuria is mild to moderate (<3 g/d) and the urine sediment inactive. In many cases of presumed hypertensive nephrosclerosis, renovascular disease may not only be the cause of hypertension but also may cause ischemic renal damage. Bilateral renovascular ischemic disease may be a greatly underdiagnosed cause of CRD. This is of therapeutic significance from two points of view: (1)

documentation of ischemic renal disease may prompt revascularization therapy in some patients, with occasional dramatic stabilization or improvement in renal function; and (2) renovascular ischemic disease is a contraindication toACEinhibitor therapy in most cases. Analgesic-associated chronic tubulointerstitial nephropathy is also an underdiagnosed cause of CRD. Imaging studies, including computed tomography, often reveal pathognomonic features such as papillary calcification and necrosis. Under such circumstances, cessation of analgesic exposure may dramatically stabilize renal function.

Kidney Biopsy This procedure should be reserved for patients with near-normal kidney size, in whom a clear-cut diagnosis cannot be made by less invasive means, and when the possibility of a reversible underlying disease process remains tenable so that clarification of the underlying etiology may alter management. The extent of tubulointerstitial scarring on kidney biopsy generally provides the most reliable pathologic correlate indicating prognosis for continued deterioration towardESRD. Contraindications to renal biopsy include bilateral small kidneys, polycystic kidney disease, uncontrolled hypertension, urinary tract or perinephric infection, bleeding diathesis, respiratory distress, and morbid obesity.

TREATMENT

This refers to all of the preventive and therapeutic measures that precede and aim to prevent or postponeESRD and renal replacement therapy.

Specific Therapy The optimal time for specific therapy aimed at the underlying disease process is usually well before there has been a measurable decline in baselineGFR, and usually well beforeCRD is established. When kidney size remains well preserved, renal biopsy results may provide an index of chronicity versus disease activity, which might help in guiding therapeutic decisions. In contrast, by the time CRD is established and GFR has irreversibly declined to less than 20 to 30% of normal, the risks of immunomodulatory and other therapies aimed at treating an underlying past or ongoing disease process may outweigh the benefits.

Superimposed Factors It is of benefit to follow and plot the rate of decline inGFR in patients withCRD. Any acceleration in the rate of decline should prompt a search for a superimposed acute process. The differential diagnosis should be developed in a systematic manner, as for any patient with acute renal failure (Chap. 269). Particular attention should be directed to factors that more commonly lead to an acute and reversible decline in GFR in patients with CRD. These include superimposed volume depletion, accelerated and uncontrolled hypertension, urinary tract infection, superimposed obstructive uropathy (e.g., due to stone disease, papillary necrosis), nephrotoxic effect of medications (e.g.,NSAIDs) and radiocontrast agents, and reactivation or flare of the original underlying etiologic disease process.

Measures to Mitigate Hyperfiltration Injury The two major therapeutic tools currently available in the mitigation of hyperfiltration injury are: (1) dietary protein restriction, and (2) pharmacologic management of intraglomerular hypertension.

Protein Restriction inCRDManagement guidelines for protein restriction in CRD are

shown in theGuidelines. In contrast to fat and carbohydrates, protein in excess of the daily requirement is not stored but is degraded to form urea and other nitrogenous wastes, which are principally excreted by the kidney. In addition, protein-rich foods contain hydrogen ions, PO_4^{3-}, sulfates, and other inorganic ions that are also eliminated by the kidney. Therefore, when patients with CRD consume excessive dietary protein, nitrogenous wastes and inorganic ions accumulate, resulting in the clinical and metabolic disturbances characteristic of uremia. Restricting dietary protein can ameliorate many uremic symptoms and may slow the actual rate of nephron injury. The effectiveness of protein restriction in slowing the progression of CRD has been evaluated in a number of controlled clinical trials. The Modification of Diet in Renal Disease (MDRD) Study was the most extensive trial devoted to this question, but it nevertheless yielded an ambiguous result, although positive trends emerged when it ended after an average follow-up of only 2.2 years. In a separate study of patients with insulin-dependent diabetic nephropathy, protein restriction was shown to slow progression significantly in one well-controlled study. Two meta-analyses of studies of the effects of protein restriction on progression concluded that low-protein diets slow progression of both diabetic and nondiabetic renal disease.

It is crucial that protein restriction be carried out in the context of an overall dietary program that optimizes nutritional status and avoids malnutrition, especially as patients near dialysis or transplantation. Measurements of urinary nitrogen appearance, anthropometric and biochemical measurements, as well as dietary consultation are mandatory to preempt malnutrition. Among the most readily available and useful indices of malnutrition are plasma concentrations of albumin (<3.8 g/dL), pre-albumin (<18 mg/dL), and transferrin (<180 ug/dL). Metabolic and nutritional studies indicate that protein requirements for patients withCRD are similar to those for normal adults, approximately 0.6 g/kg per day. However, there is a particular requirement in patients with CRD that the composition of dietary protein be higher in essential amino acids, and that this be combined with an overall energy supply sufficient to mitigate a catabolic state. Energy requirements in the range of 35 kcal/kg per day are recommended.

Fortunately, even patients with advancedCRD are able to activate the same adaptive responses to dietary protein restriction as healthy individuals, i.e., a postprandial suppression of whole-body protein degradation and a marked inhibition of amino acid oxidation. After at least 1 year of therapy with a low-protein diet (range 12 to 24 months) these same adaptive responses persist, indicating that the compensatory responses to dietary protein restriction are sustained during long-term therapy. Further evidence that low-protein diets are safe in CRD patients is provided by the finding that nutritional indices remain normal during long-term therapy.

Pharmacologic Management of Intraglomerular Hypertension In addition to reduction of cardiovascular disease risk, antihypertensive therapy in patients withCRD also aims to slow the progression of nephron injury by ameliorating intraglomerular hypertension and hypertrophy. Progressive renal injury in CRD appears to be most closely related to the height of intraglomerular pressure and/or the extent of glomerular hypertrophy. TheMDRD and other studies demonstrated that control of hypertension is as important as dietary protein restriction in slowing the progression of CRD. Furthermore, the target for pharmacologic therapy was highly dependent on the level of proteinuria. Indeed, proteinuria is now considered a risk factor for progressive nephron

injury; the prior level of proteinuria correlates with the subsequent rate ofGFRdecline. Elevated blood pressure increases proteinuria due to the transmission to the glomeruli of the elevated systemic pressure. Conversely, the protective effect of antihypertensive medications is evident through the curtailment of proteinuria. Thus, the more effective a given treatment is in lowering proteinuria, the greater the subsequent impact on protection from GFR decline.

Some antihypertensive agents, particularly theACEinhibitors, may be superior to others in affording renal protection. The advantage of this pharmacologic class is thought to relate to salutary modulation of intraglomerular hemodynamics over and above effects on systemic blood pressure. Several well-designed studies have now established favorable outcomes for ACE inhibitors in slowing the progression of diabetic nephropathy. ACE inhibitors have been shown to be more effective than diuretics, beta blockers, and calcium antagonists in reducing urinary albumin excretion in both hypertensive and diabetic patients. Furthermore, these drugs have been shown to facilitate the regression of remodeling more generally in the cardiovascular system and to improve endothelial function in resistance arterioles of humans with hypertension. In nondiabeticCRD, the European AIPRI trial documented a 53% additional reduction in the risk of doubling serum creatinine levels with ACE inhibitor therapy compared to conventional regimens that did not include ACE inhibitors. The reduction in the risk was greater in patients with mild renal insufficiency and in those with proteinuria >1g/d. In a recent meta-analysis, information from all the randomized ACE inhibitor trials in patients with nondiabetic renal disease was combined; the conclusion is that ACE inhibitors are more effective than other antihypertensive agents in reducing the development ofESRD. A similar salutary effect on kidney function as observed with ACE inhibitors has been observed recently with the angiotensin II receptor antagonists, which also possess significant antiproteinuric properties.

Among the calcium channel blockers, diltiazem and verapamil appear to exhibit antiproteinuric and renal protective effects not shared by the dihydropyridines. As a group, these drugs do not adversely affect renal function in patients with nondiabetic renal insufficiency, and they may be more effective in preventing or ameliorating progressive renal injury than some other classes of antihypertensive drugs in this group of patients. Thus, it appears that at least two different categories of responses may exist: one in which progression is strongly associated with systemic and intraglomerular hypertension and with proteinuria (e.g., diabetic nephropathy, glomerular diseases) and in whichACEinhibitors and angiotensin-receptor blockers are likely to be the first choice; and the second in which proteinuria is mild or absent (e.g., adult polycystic kidney disease), probably with a less prominent role for intraglomerular hypertension, and which might respond as well to calcium channel blockers. The level of blood pressure lowering is also of crucial importance in achieving a significant renal protective effect. Clinical practice guidelines are summarized in theGuidelines.

Use of Drugs (See also Chaps. 70 and71) Although the loading dose of most drugs is not affected byCRD, maintenance doses of many drugs need to be adjusted. One exception is digoxin, whose volume of distribution is decreased in CRD, mandating a concomitant reduction in the loading dose in addition to adjustment of the maintenance dose. For those drugs in which >70% excretion is by a nonrenal (e.g., hepatic or intestinal) route, dosage adjustment may not be needed. Some drugs that should be

entirely avoided include meperidine, metformin, and other oral hypoglycemics with a renal route of elimination. Commonly used medications that require either a reduction in dosage or interval include allopurinol, many antibiotics, several hypertensives, and anti-arrhythmics. For a comprehensive detailed and authoritative listing of the recommended dose adjustment for most of the commonly used medications, the reader is referred to the American College of Physicians handbook of "Drug Prescribing in Renal Failure" (seehttp://www.acponline.org).

Preparation for Renal Replacement Therapy Over the past 35 years, renal replacement therapy using dialysis and transplantation has prolonged the lives of hundreds of thousands of patients withESRD. Renal replacement therapy should *not* be initiated when the patient is totally asymptomatic; however, dialysis and/or transplantation should be started sufficiently early to prevent serious complications of the uremic state. Clear indications for initiation of renal replacement therapy include pericarditis, progressive neuropathy attributable to uremia, encephalopathy, muscle irritability, anorexia and nausea that is notameliorated by reasonable protein restriction, and fluid and electrolyte abnormalities that are refractory to conservative measures. The latter include volume overload unresponsive to diuretic therapy, hyperkalemia unresponsive to dietary potassium restriction, and progressive metabolic acidosis that cannot be managed with alkali therapy. Clinical clues indicating the imminent development of uremic complications are a history of hiccoughing, intractable pruritus, morning nausea and vomiting, muscle twitching and cramps, and the presence of asterixis on physical examination. In addition, the patient whose follow-up and compliance with conservative management are questionable should be considered for earlier initiation of renal replacement therapy, lest potentially life-threatening uremic complications or electrolyte disturbances supervene.

The correlation of uremic symptoms with renal function varies from patient to patient depending on the cause of renal disease (earlier onset of symptoms in patients with diabetes mellitus), muscle mass (large, muscular patients tolerate high levels of azotemia), diet, nutritional status, and coexisting conditions. Therefore, it is ill-advised to assign a certain "usual" level of blood urea nitrogen, serum creatinine, or GFR to the need to start dialysis. Nevertheless, in the United States, the Health Care Financing Administration has assigned levels of serum creatinine and creatinine clearance to qualify for reimbursement from Medicare for patients receiving dialysis. Serum creatinine must be ³700 umol/L (³8.0 mg/dL) and the creatinine clearance must be £0.17 mL/s (£10 mL/min).

Patient Education Social, psychological, and physical preparation for the transition to renal replacement therapy and choice of the optimal initial modality is best accomplished with a gradual approach involving a multidisciplinary team. While conservative measures are being carried out in patients withCRD, it is important to prepare them with an intensive educational program, explaining the likelihood and timing of initiation of renal replacement therapy and the various forms of therapy available. The more knowledgeable patients are concerning hemodialysis, peritoneal dialysis, and transplantation, the easier and more appropriate will be their decisions at a later time. Exploration of social service support resources is of great importance. In those who may perform home dialysis or undergo transplantation, early education of family members for selection and preparation as a home dialysis helper or a related

donor for transplantation should occur long before the onset of symptomatic renal failure.

Selection of patients to be treated with various modalities of dialysis or transplantation is a matter of some debate, with considerable variation in different parts of the world. In general, in the United States and some other countries, nearly all patients who have reachedESRD are accepted for dialysis if they or their families desire prolongation of life, irrespective of age.

In terms of dialysis treatment modalities (Chap. 271), large multicenter studies have not shown a consistent or convincing advantage in terms of morbidity or mortality, of one modality over another.

Only kidney transplantation (Chap. 272) offers the potential for nearly complete rehabilitation. This is because dialysis techniques replace only 10 to 15% of normal kidney function at the level of small-solute removal and are even less efficient at the removal of larger solutes. Generally, kidney transplantation follows a prior period of dialysis treatment. All patients in whom an acute reversible component of renal failure has not been completely excluded should be supported with dialysis first, at least for some period of time, to allow for possible return of renal function before consideration of transplantation. Recovery of endogenous renal function in patients treated with dialysis for more than 6 months is a rare occurrence. Usually these are patients in whom the underlying disease process has been acute or subacute -- such as one of the thrombotic microangiopathies, rapidly progressive glomerulonephritis, or obstructive uropathy. Patients approachingESRD in whom a reversible component has been excluded, and who have a good antigenic match with a willing donor, may occasionally be considered for primary transplantation without intervening dialysis.

ACKNOWLEDGEMENT

Dr. J. Michael Lazarus was a co-author of thischapterin the 14th edition, and some of the material in thatchapteris carried forward to the present edition.

(Bibliography omitted in Palm version)

Back to Table of Contents

271. DIALYSIS IN THE TREATMENT OF RENAL FAILURE - *Ajay K. Singh, Barry M. Brenner*

With the widespread availability of dialysis, the lives of hundreds of thousands of patients with end-stage renal disease (ESRD) have been prolonged. In the United States alone, there are now approximately 300,000 patients with ESRD. The overall incidence of ESRD is 242 cases per million population per year. The incident population of patients with ESRD is increasing at approximately 8% each year. The incidence of ESRD is disproportionately higher in African Americans (758 per million population per year) as compared with white Americans (180 per million population per year). In the United States, the leading cause of ESRD is diabetes mellitus, accounting for more than 40% of newly diagnosed cases of ESRD. The second most common cause is hypertension, which is estimated to cause 30% of ESRD cases. Other causes of ESRD include glomerulonephritis, polycystic kidney disease, and obstructive uropathy.

Dialysis care in the United States is funded through the Medicare End-Stage Renal Dialysis program; the cost was approximately $13.5 billion in 1997. Although the total program expense has increased dramatically, the cost for treating individual patients in inflation-adjusted terms has gone down. In addition, quality and outcomes have improved through better management of dialysis dose, improved nutrition, management of anemia, and control of hypertension. The cost of dialysis, which ranges from $45,000 to $65,000 per year, has been demonstrated to vary according to the presence or absence of diabetes mellitus and whether the treatment modality is hemodialysis or peritoneal dialysis. The mortality of patients withESRD is lowest in Europe and Japan but is very high in the developing world because of the limited availability of dialysis. In the United States, the mortality rate of patients on dialysis is approximately 18% per year. Deaths are due mainly to cardiovascular diseases and infections (approximately 50% and 15% of deaths, respectively).

TREATMENT OPTIONS FOR ESRD PATIENTS

Commonly accepted criteria for putting patients on dialysis include: the presence of the uremic syndrome; the presence of hyperkalemia unresponsive to conservative measures; extracellular volume expansion; acidosis refractory to medical therapy; a bleeding diathesis; and a creatinine clearance of<10 cc/min per 1.73m_2. There is emerging consensus that patients with ESRD should be started on dialysis early. Although vigorous protein restriction can maintain the blood urea nitrogen at an acceptable level in these patients, it may come at the price of significant malnutrition, which in turn correlates with mortality on dialysis. In addition to carefully evaluating patients for the onset of uremia (Chap. 270), regular measurement of renal function is important.

Renal function can be assessed by measurement of serum creatinine and blood urea nitrogen or of creatinine and urea clearance, or the direct measurement of glomerular filtration rate (GFR) using a radioisotope such as iothalamate. Creatinine clearance usually overestimates glomerular filtration rate because a substantial fraction of creatinine excretion in advanced renal failure occurs as a consequence of proximal tubular secretion. On the other hand, urea clearance invariably underestimates GFR because urea is reabsorbed in the distal nephron. Thus, when measurement of GFR by

a direct test is not available, the average of the sum of the creatinine and urea clearance, or a cimetidine-blocked creatinine clearance (cimetidine blocks proximal tubular secretion), is recommended. Early referral to a nephrologist for advanced planning and creation of a dialysis access, education aboutESRDtreatment options, and the aggressive management of the complications of chronic renal failure, including acidosis, anemia, and hyperparathyroidism, are important.

The treatment options available for patients with renal failure depend on whether it is acute or chronic (Fig. 271-1). In acute renal failure, treatments include hemodialysis, continuous renal replacement therapies (see p. 1565), and peritoneal dialysis. In chronic renal failure(ESRD)the options include hemodialysis (in center or at home); peritoneal dialysis, either as continuous ambulatory peritoneal dialysis (CAPD) or continuous cyclic peritoneal dialysis (CCPD); or transplantation (Chap. 272). Although there are geographic variations, hemodialysis remains the most common therapeutic modality for ESRD (>80% of patients in the United States). The choice between hemodialysis and peritoneal dialysis involves the interplay of various factors that include the patient's age, the presence of comorbid conditions, the ability to perform the procedure, and the patient's own conceptions about the therapy. Peritoneal dialysis is favored in younger patients because of their better manual dexterity and greater visual acuity, and because younger patients prefer the independence and flexibility of home-based peritoneal dialysis treatment. In contrast, larger patients (>80 kg), patients with no residual renal function, and patients who have truncal obesity with or without prior abdominal surgery are more suited to hemodialysis. Larger patients with no residual renal function are more appropriate for hemodialysis because these patients have a large volume of distribution of urea and require significantly higher amounts of peritoneal dialysis, which may be difficult to achieve because of the limited willingness of patients to perform more than four exchanges each day. In some patients, the inability to obtain vascular access predicates a switch from hemodialysis to peritoneal dialysis.

HEMODIALYSIS

This consists of diffusion that occurs bi-directionally across a semipermeable membrane. Movement of metabolic waste products takes place down a concentration gradient from the circulation into the dialysate, and in the reverse direction. The rate of diffusive transport increases in response to several factors, including the magnitude of the concentration gradient, the membrane surface area, and the mass transfer coefficient of the membrane. The latter is a function of the porosity and thickness of the membrane, the size of the solute molecule, and the conditions of flow on the two sides of the membrane. According to the laws of diffusion, the larger the molecule, the slower its rate of transfer across the membrane. A small molecule such as urea (60 Da) undergoes substantial clearance, whereas a larger molecule such as creatinine (113 Da), is cleared much less efficiently. In addition to diffusive clearance, movement of toxic materials such as urea from the circulation into the dialysate may occur as a result ofultrafiltration. Convective clearance occurs because of solvent drag with solutes getting swept along with water across the semipermeable dialysis membrane.

THE DIALYZER

There are three essential components to dialysis: the dialyzer, the composition and

delivery of the dialysate, and the blood delivery system (Fig. 271-2). The dialyzer consists of a plastic device with the facility to perfuse blood and dialysate compartments at very high flow rates. The surface area of dialysis membranes in adult patients is usually in the range of 0.8 to 1.2 m_2.

There are currently two geometric configurations for dialyzers: hollow fiber and flat plate. The hollow fiber dialyzer is the most common in use in the United States. These dialyzers are composed of bundles of capillary tubes through which blood circulates while dialysate travels on the outside of the fiber bundle. In contrast, the less frequently utilized flat plate dialyzers are composed of sandwiched sheets of membrane in a parallel plate configuration. The advantage of the hollow fiber construction is the lower priming volume (60 to 90 mL vs 100 to 120 mL for the flat plate) and easier reprocessing of the filter for reuse in future dialysis treatments.

Recent advances have led to the development of many different types of membrane material. Broadly, there are four categories of dialysis membranes: cellulose, substituted cellulose, cellulosynthetic, and synthetic. Over the past two decades, there has been a gradual switch from cellulose-derived to synthetic membranes, because the latter are more biocompatible. Bioincompatibility may be defined as the ability of the membrane to activate the complement cascade. Cellulosic membranes are bioincompatible because of the presence of free hydroxyl groups on the membrane surface. In contrast, with the substituted cellulose membranes (e.g., cellulose acetate) or the cellulosynthetic membranes, the hydroxyl groups are chemically bonded to either acetate or tertiary amino groups, resulting in limited complement activation. Synthetic membranes, such as polysulfone, polymethylmethacrylate, and polyacrylonitrile membranes are more biocompatible because of the absence of these hydroxyl groups. Polysulfone membranes are now used in over 60% of the dialysis treatments in the United States.

Reprocessing and reuse of hemodialyzers is employed for patients on chronic hemodialysis in nearly 80% of dialysis centers in the United States, in large part because of the expense of individual dialyzers. Evidence also suggests that reuse reduces complement activation, the incidence of anaphylactoid reactions to the membrane (first-use syndrome), and, in some studies, mortality rates among dialysis patients. In most centers, only the dialyzer unit is reprocessed and reused, whereas in the developing world blood lines are also frequently reused. The reprocessing procedure can be either manual or automated. It consists of the sequential rinsing of the blood and dialysate compartments with water, a chemical cleansing step with reverse ultrafiltration from the dialysate to the blood compartment, the testing of the patency of the dialyzer, and, finally, disinfection of the dialyzer. Formaldehyde, peracetic acid-hydrogen peroxide, and glutaraldehyde are the most frequently used reprocessing agents, with peracetic acid-hydrogen peroxide being the most common.

DIALYSATE

The composition of dialysate is listed inTable 271-1. Bicarbonate has replaced acetate as the preferred buffer in the United States. This change has resulted in fewer episodes of hypotension during dialysis. The potassium concentration of dialysate may be varied from 0 to 4 mmol/L depending on the predialysis plasma potassium concentration. The usual dialysate calcium concentration is 1.25 mmol/L (2.5 meq/L). The usual dialysate

sodium concentration is 140 mmol/L. Lower dialysate sodium concentrations are associated with a higher frequency of hypotension, cramping, nausea, vomiting, fatigue, and dizziness. In patients who frequently develop hypotension during their dialysis run, sodium modeling to counterbalance urea-related osmolar gradients is now widely used. In this technique, the dialysate sodium concentration is gradually lowered from the range of 148 to 160 meq/L to isotonic levels (140 meq/L) near the end of the dialysis treatment. A dialysate glucose concentration of 200 mg/dL (11 mmol/L) is used to optimize blood glucose concentrations. Because patients are exposed to approximately 120 L of water during each dialysis treatment, untreated water could expose them to a variety of environmental contaminants. Therefore, in 98% of U.S. dialysis centers, water used for the dialysate is subjected to filtration, softening, deionization, and, ultimately, reverse osmosis. During the reverse osmosis process, water is forced through a semipermeable membrane at very high pressure to remove microbiologic contaminants and more than 90% of dissolved ions.

BLOOD DELIVERY SYSTEM

This is composed of the extracorporeal circuit in the dialysis machine and the dialysis access. The dialysis machine consists of a blood pump, dialysis solution delivery system, and various safety monitors. The blood pump, using a roller mechanism, moves blood from the access site, through the dialyzer, and back to the patient. The blood flow rate may range from 250 to 500 mL/min. Negative hydrostatic pressure on the dialysate side can be manipulated to achieve desirable fluid removal, so-called *ultrafiltration*. Dialysis membranes have different ultrafiltration coefficients (i.e., mL removed/min per mmHg) so that along with hydrostatic changes, fluid removal can be varied. The dialysis solution delivery system dilutes the dialysate concentrate with water, and monitors the temperature, conductivity, and flow of dialysate. The dialysate may be delivered to the dialyzer from a storage tank or a proportioning system that manufactures dialysate online.

Dialysis Access The fistula, graft, or catheter through which blood is obtained for hemodialysis is often referred to as a *dialysis access*. A native fistula created by the anastomosis of an artery to a vein (e.g., the Cimino-Breschia fistula, in which the cephalic vein is anastomosed to the radial artery) results in arterialization of the vein. This faciliates its subsequent use in the placement of large needles (typically 15 gauge) to access the circulation. Although fistulas have a high patency rate (approximately 80% are patent at 3 years following creation), fistulas are created in only approximately 30% of patients in the United States. In the majority of U.S. dialysis patients, the dialysis access consists of an arteriovenous graft which interposes prosthetic material, such as polytetrafluoroethylene, between an artery and a vein. Such grafts have a 3-year patency rate of only 20%. Reasons for the higher rates of graft placement include the late referral of patients to vascular access surgeons so that by the time surgery is planned, the patient's arm veins have already been obliterated through multiple blood draws; the high prevalence of patients with diabetes mellitus and its associated microvascular disease; and the greater surgical skill required in creating a fistula. The most common access-related complication is thrombosis due to intimal hyperplasia, which results in stenosis proximal to the venous anastomosis.

A double lumen cuffed catheter may be a reasonable alternative to either a native

arteriovenous fistula or a graft in selected patients in whom dialysis is required relatively urgently, such as patients who manifest delayed recovery from acute renal failure, or where a further permanent access procedure (e.g., arteriovenous fistula or arteriovenous graft) is not feasible for anatomic reasons. Although double lumen catheters may permit blood flows comparable to a permanent arteriovenous access, these catheters are prone to infection and to occlusion because of thrombosis. Temporary double lumen catheters in either the femoral vein or the internal jugular or subclavian vein are usually employed in patients with acute renal failure. The jugular is preferred to the subclavian vein because, for unclear reasons, a catheter placed in a subclavian vein appears to be associated with a higher rate of venous stenosis. Temporary access can be used for 2 to 3 weeks. Thromobosis, low blood flow, and infection limit the life of the catheter.

GOALS OF DIALYSIS

The hemodialysis procedure is targeted at removing both small and large molecular weight solutes. The procedure consists of pumping heparinized blood through the dialyzer at a flow rate of 300 to 500 mL/min, while dialysate flows in an opposite *counter-current* direction at 500 to 800 mL/min. The clearance of urea ranges from 200 to 350 mL/min, while the clearance of$b_2$microglobulin is more modest and ranges from 20 to 25 mL/min. The efficiency of dialysis is determined by blood and dialysate flow through the dialyzer, as well as dialyzer characteristics (i.e., its efficiency in removing solute). The *dose* of dialysis, which is defined as the magnitude of urea clearance during a single dialysis treatment, is further governed by patient size, residual renal function, dietary protein intake, the degree of anabolism or catabolism, and the presence of comorbid conditions. Since the landmark studies of Sargent and Gatch relating the measurement of the dose of dialysis using urea concentration with patient outcome, the *delivered* dose of dialysis has been correlated with morbidity and mortality. This has led to the development of two major models for assessing the adequacy of the dialysis dose. Fundamentally, these two widely used measures of the adequacy of dialysis are calculated from the decrease in the blood urea nitrogen concentration during the dialysis treatment -- that is, the urea reduction ratio (URR), and *KT/V*, an index based on the urea clearance rate, *K*, and the size of the urea pool, represented as the urea distribution volume, *V*. K, which is the sum of clearance by the dialyzer plus renal clearance, is multiplied by the time spent on dialysis, *T*. Increasingly, KT/V has become the preferred marker for dialysis adequacy. Currently, a URR of 65% and a KT/V of 1.2 per treatment are minimal standards for adequacy; lower levels of dialysis treatment are associated with increased morbidity and mortality.

For the majority of patients with chronic renal failure, between 9 and 12 h of dialysis is required each week, usually divided into three equal sessions. However, the dialysis dose must be individualized. The measurement of dialysis adequacy using KT/V or theURRserve only as a guide; body size, residual renal function, dietary intake, complicating illness, degree of anabolism or catabolism, and the presence of large interdialytic fluid gains are important factors in consideration of the dialysis prescription.

COMPLICATIONS DURING HEMODIALYSIS

Hypotension is the most common acute complication of hemodialysis. Numerous factors

appear to increase the risk of hypotension, including excessive ultrafiltration with inadequate compensatory vascular filling, impaired vasoactive or autonomic responses, osmolar shifts, food ingestion, impaired cardiac reserve, the use of antihypertensive drugs, and vasodilation due to the use of warm dialysate. Because of the vasodilatory and cardiodepressive effects of acetate, the use of acetate as the buffer in dialysate was once a common cause of hypotension. Since the introduction of bicarbonate-containing dialysate, dialysis-associated hypotension has become common. The management of hypotension during dialysis consists of discontinuing ultrafiltration, the administration of 100 to 250 cc of isotonic saline, and, in patients with hypoalbuminemia, administration of salt-poor albumin. Hypotension during dialysis can frequently be prevented by careful evaluation of the dry weight, holding of antihypertensive medications on the day prior to and on the day of dialysis, and avoiding heavy meals during dialysis. Additional maneuvers include the performance of sequential ultrafiltration followed by dialysis and cooling of the dialysate during dialysis treatment.

Muscle cramps during dialysis are also a common complication of the procedure. However, since the introduction of volumetric controls on dialysis machines and sodium modelling, the incidence of cramps has fallen. The etiology of dialysis-associated cramps remains obscure. Changes in muscle perfusion because of excessively aggressive volume removal, particularly below the estimated dry weight and the use of low sodium containing dialysate, have been proposed as precipitants of dialysis-associated cramps. Strategies that may be used to prevent cramps include reducing volume removal during dialysis, the use of higher concentrations of sodium in the dialysate, and the use of quinine sulfate (260 mg 2 h before treatment).

Anaphylactoid reactions to the dialyzer, particularly on its first use, have been reported most frequently with the bioincompatible cellulosic-containing membranes. With the gradual phasing out of cuprophane membranes in the United States, the first use syndrome has become relatively uncommon. The first use syndrome consists of either an intermediate hypersensitivity reaction due to an IgE mediated reaction to ethylene oxide used in the sterilization of new dialyzers, or a symptom complex of nonspecific chest and back pain, which appears to result from complement activation and cytokine release.

The major cause of death in patients withESRDreceiving chronic dialysis is cardiovascular disease. The rate of death from cardiac disease is higher in patients on hemodialysis as compared to patients on peritoneal dialysis and renal transplantation. The underlying cause of cardiovascular disease is unclear but may be related to the inadequate treatment of hypertension; the presence of hyperlipidemia, homocystinemia and anemia; the calcification of coronary arteries in patients with an elevated calcium-phosphorus product; and perhaps alterations in cardiovascular dynamics during the dialysis treatment. Intensive investigation of the mechanisms and potential interventions that could impact on reducing the mortality from cardiovascular causes is currently underway.

CONTINUOUS RENAL REPLACEMENT THERAPY

Continuous renal replacement therapies (CRRT) have become increasingly prevalent in

the intensive care unit setting for management of acute renal failure. The advantages of CRRT over intermittent hemodialysis are that it is usually better tolerated hemodynamically; it facilitates gradual correction of biochemical abnormalities; it is highly effective in removing fluid; and it is technically simple to perform. Clearance of toxic materials (using urea as the marker) can occur with CRRT from convective clearance alone if the ultrafiltration rate is high and with diffusive clearance if dialysis accompanies ultrafiltration. CRRT techniques include continuous arteriovenous hemodiafiltration (CAVH/D) with or without dialysis, and continuous veno-venous hemodiafiltration (CVVH/D) with or without dialysis. Veno-venous therapies differ fundamentally from arteriovenous therapies in that veno-venous therapies do not require arterial access. This allows obtaining less risky and easier vascular access. However, because there is no systemic arterial pressure to drive hemofiltration, veno-venous therapies require a blood pump in the extracorporeal circuit. Veno-venous therapies such as CVVH provide substantial flexibility because changing the blood flow rate in the pump can change the ultrafiltration and clearance rates. In contrast, arterio-venous therapies such as CAVH are associated with variable efficiency because the systemic blood pressure is frequently low or unstable in patients with acute renal failure. Furthermore, low blood flow with CAVH may also result in clotting of the extracorporeal circuit. CAVH often results in clearance rates as low as 10 to 15 mL/min, whereas CVVH may generate clearances in the range of 30 to 40 mL/min. Thus, in light of these advantages of CVVH, many centers have completely switched from arteriovenous to veno-venous therapies in patients with acute renal failure in the ICU setting.

Vascular access in patients onCVVH is usually achieved by the insertion of a double-lumen catheter into the femoral vein. The blood pump is typically set to deliver approximately 150 to 180 mL/min. In automated systems, (e.g., the Cobe Prisma system), the treatment is volumetrically governed by continuously weighing the effluent and replacement solutions and using a servomechanism to drive the replacement fluid pump at a rate computed either to balance the inflow and loss of fluid or to maintain a predetermined rate of fluid loss. Anticoagulation of the extracorporeal circuit is via a heparin infusion (200 to 1600 U/h) through the inflow side of the circuit. Alternatively, citrate can be used to chelate calcium in the extracorporeal circuit to provide regional anticoagulation in selected patients who cannot undergo systemic heparinization. The replacement solution in continuous therapies is designed specifically to replace calcium, magnesium, and bicarbonate. In place of bicarbonate, lactate or citrate is the buffer in the replacement solution. However, bicarbonate-based replacement fluid is the preferred option in patients with liver failure because of the impaired ability of the liver to metabolize either lactate or acetate into bicarbonate.

PERITONEAL DIALYSIS

This consists of infusing 1 to 3 L of a dextrose-containing solution into the peritoneal cavity and allowing the fluid to dwell for 2 to 4 h. As with hemodialysis, toxic materials are removed through a combination of convective clearance generated through ultrafiltration, and diffusive clearance down a concentration gradient. The clearance of solute and water during a peritoneal dialysis exchange depends on the balance between the movement of solute and water into the peritoneal cavity versus absorption from the peritoneal cavity. The rate of diffusion diminishes with time and eventually

stops when equilibriation between plasma and dialysate is reached. Absorption of solutes and water from the peritoneal cavity occurs across the peritoneal membrane into the peritoneal capillary circulation and via peritoneal lymphatics into the lymphatic circulation. The rate of peritoneal solute transport varies from patient to patient and may be altered by the presence of infection (peritonitis), drugs such as beta blockers and calcium channel blockers, and by physical factors such as position and exercise.

FORMS OF PERITONEAL DIALYSIS

Peritoneal dialysis may be carried out as continuous ambulatory peritoneal dialysis (CAPD),continuous cyclic peritoneal dialysis (CCPD), or nocturnal intermittent peritoneal dialysis (NIPD). In CAPD, dialysis solution is manually infused into the peritoneal cavity during the day and exchanged 3 to 4 times daily. A nighttime dwell is frequently instilled at bedtime and remains in the peritoneal cavity through the night. The drainage of spent dialysate (effluence) is performed manually with the assistance of gravity to move fluid out of the abdomen. In CCPD, exchanges are performed in an automated fashion, usually at night; the patient is connected to the automated cycler, which then performs 4 to 5 exchange cycles while the patient sleeps. Peritoneal dialysis cyclers automatically cycle dialysate in and out of the abdominal cavity. In the morning the patient, with the last exchange remaining in the abdomen, is disconnected from the cycler and goes about his regular daily activities. In NIPD, the patient is given approximately 10 h of cycling each night, with the abdomen left dry during the day.

Peritoneal dialysis solutions are available in various volumes ranging from 0.5 to 3.0 L. The electrolyte composition is shown inTable 271-2. Lactate is the preferred buffer in peritoneal dialysis solutions. Acetate in peritoneal dialysis solutions appears to accelerate peritoneal sclerosis, whereas use of bicarbonte results in precipitation of calcium and caramelization of glucose. The most common additives to peritoneal dialysis solutions are heparin and antibiotics during an episode of acute peritonitis. Insulin may also be added in patients with diabetes mellitus.

ACCESS TO THE PERITONEAL CAVITY

This is obtained through a peritoneal catheter. These are either *acute* catheters, used to perform acute continuous peritoneal dialysis, usually in an emergency setting, or *chronic* catheters, which have either one or two Dacron cuffs and are tunneled under the skin into the peritoneal cavity. An acute catheter consists of a straight or slightly curved rigid tube with several holes at its distal end. Catheters can be inserted at the bedside by making a small incision in the anterior abdominal wall; the catheter is inserted with the assistance of a guidewire or stylet. Acute catheters are anchored externally with adhesives or sutures and are usually reserved for temporary use because of the risk of infection, which increases after 72 h of use. In contrast, chronic catheters are flexible and made of silicon rubber with numerous side holes at the distal end. These chronic catheters usually have two Dacron cuffs to promote fibroblast proliferation, granulation and invasion of the cuff. The scarring that occurs around the cuffs anchors the catheter and seals it from bacteria tracking from the skin surface into the peritoneal cavity; it also prevents the external leakage of fluid from the peritoneal cavity. The cuffs are placed in the preperitoneal plane and approximately 2 cm from the skin surface. The most common chronic peritoneal dialysis catheter in use is the Tenckhoff catheter, which

contains two cuffs.

The initialCAPDprescription consists of the infusion of a 2-L volume of a 1.5% dextrose concentration peritoneal dialysis solution into the peritoneal cavity over 10 min and allowing it to dwell for 2.5 h. The effluent solution is then drained over 20 min before the next exchange. Three daytime exchanges are accompanied by a 2 L nighttime dwell as the standard prescription. Because peritoneal membrane characteristics vary from one individual to another, the peritoneal equilibrium test should be employed within 2 months of a patient initiating peritoneal dialysis. This test measures the peritoneal membrane transfer rate for solutes (usually urea and creatinine) based on the ratio of their concentration in dialysate and plasma at specific times during the dialysate dwell. It allows patients to be classified as low, low-average, high-average, and high transporters. Approximately 10 to 17% of patients are high transporters, 50% high-average transporters, 25 to 30% low-average transporters, and 1 to 5% low transporters. Identifying the high transporters early is important, since these patients not only demonstrate excellent solute removal, they also absorb glucose rapidly; maximum ultrafiltration occurs early in the dwell, followed by reabsorption of water back into the circulation over the course of the dwell. Such patients benefit from eitherNIPD or CAPD without a nighttime dwell.

The dose of peritoneal dialysis required to provide adequate or optimal dialysis as measured by patient outcomes is not known. However, there is emerging consensus that the weekly KT/V should be >2.0 and the creatinine clearance >65 L/week per 1.73 m_2. The most frequently utilized approach to calculating a weekly KT/V and creatinine clearance is by collecting the spent dialysate and urine over a 24-h period. The peritoneal dialysis prescription can be tailored to improve suboptimal clearance values by either increasing the volume of individual exchanges, increasing the number of exchanges, or by combining theCAPD andCCPDtechniques. In combining these techniques, the CAPD patient hooks up to a cycler at night and the machine automatically performs one or two nocturnal exchanges, whereas the CCPD patient makes an additional manual daytime exchange.

ACKNOWLEDGEMENT

Dr. J. Michael Lazarus was a co-author of thischapterin the 14th edition; some of his material has been carried forward to the present edition.

(Bibliography omitted in Palm version)

Back to Table of Contents

272. TRANSPLANTATION IN THE TREATMENT OF RENAL FAILURE - *Charles B. Carpenter*, *Edgar L. Milford*, *Mohamed H. Sayegh*

Transplantation of the human kidney is frequently the most effective treatment of advanced chronic renal failure. Worldwide, tens of thousands of such procedures have been performed. When azathioprine and prednisone were initially used as immunosuppressive drugs in the 1960s, the results with properly matched familial donors were superior to those with organs from cadaveric donors, namely, 75 to 90% compared with 50 to 60% graft survival rates at 1 year. During the 1970s and 1980s, the success rate at the 1-year mark for cadaveric transplants rose progressively. By the time cyclosporine was introduced in the early 1980s, cadaveric donor grafts had a 70% 1-year survival and reached the 80 to 85% level in the mid 1990s (Fig. 272-1). After the first year, graft survival curves show an exponential decline in numbers of functioning grafts from which a half-life ($t_{1/2}$) in years is calculated (Fig. 272-1). Mortality rates after transplantation are highest in the first year and are age-related: 2% for ages 6 to 45 years, 7% for ages 46 to 60 years, and 10% for ages over 60 years, and lower thereafter. These rates compare favorably to those in the chronic dialysis population, even after risk adjustments for age, diabetes, and cardiovascular status. Occasionally, acute irreversible rejection may occur after many months of good function, especially if the patient neglects to take the immunosuppressive drugs. Most grafts, however, succumb at varying rates to a chronic vascular and interstitial obliterative process termed *chronic rejection*, although its pathogenesis is incompletely understood. Overall, transplantation returns the majority of patients to an improved life-style and an improved life expectancy, as compared to patients on dialysis; however, careful prospective cohort studies have yet to be reported.

RECIPIENT SELECTION

Transplantation should be undertaken only when there is a state of irreversible renal failure. When a living donor is available, a period of chronic dialysis may be avoided. When end-stage renal disease is the result of diabetes mellitus, there is special merit in having a transplant before it is necessary to initiate maintenance dialysis in order to minimize progression of cardiovascular complications of diabetes, which are frequently accelerated during chronic dialysis. In patients who must wait for a cadaveric donor kidney, a dialysis program must be established since the waiting time will be in the 3 to 4 year range for most patients. Each candidate must have a careful risk/benefit evaluation. Elderly patients over age 70, patients with metastatic malignancy, or those with advanced cardiopulmonary disease are generally poor operative risks and are also more susceptible to infections in the setting of immunosuppressive medications. Coronary artery revascularization may be indicated in select patients prior to transplantation. Because of the growing shortage of available cadaveric organs in relation to the expanding chronic dialysis population, patients who do not have a life expectancy of at least 5 years are generally not placed on the national waiting list in the United States.

DONOR SELECTION

Donors can be cadavers or volunteer living donors. The latter are usually family members selected to have at least partial compatibility for HLA antigens. Living

volunteer donors should be normal on physical examination and of the same major ABO blood group, because crossing major blood group barriers prejudices survival of the allograft. It is possible, however, to transplant a kidney of a type O donor into an A, B, or AB recipient. Selective renal arteriography should be performed on donors to rule out the presence of multiple or abnormal renal arteries, because the surgical procedure is difficult and the ischemic time of the transplanted kidney long when vascular abnormalities exist. Cadaveric donors should be free of malignant neoplastic disease, hepatitis, and HIV because of possible transmission to the recipient. Increased risk of graft failure exists when the donor is elderly or has renal failure and when the kidney has a prolonged period of ischemia and storage.

In the United States, there is a coordinated national system (United Network for Organ Sharing) of regulations, allocation support, and outcomes analysis for kidney transplantation. It is now possible to remove cadaver kidneys and to maintain them for up to 48 h on cold pulsatile perfusion or simple flushing and cooling. This permits adequate time for typing, cross-matching, transportation, and selection problems to be solved.

TISSUE TYPING AND CLINICAL IMMUNOGENETICS

Matching for antigens of the HLA major histocompatibility gene complex (Chap. 306) is an important criterion for selection of donors for renal allografts. Each mammalian species has a single chromosomal region that encodes the strong, or major, transplantation antigens, and this region on the human sixth chromosome is called *HLA*. HLA antigens have been classically defined by serologic techniques, but methods to define specific nucleotide sequences in genomic DNA are increasingly being used. Other antigens, called "minor," may nevertheless play crucial roles, in addition to the ABH(O) blood groups and endothelial antigens that are not shared with lymphocytes. The Rh system is not expressed on graft tissue. Evidence for designation of HLA as the genetic region encoding major transplantation antigens comes from the success rate in living related donor renal and bone marrow transplantation, with superior results in HLA-identical sibling pairs. Nevertheless, 5% of HLA-identical renal allografts are rejected, often within the first weeks after transplantation. These failures represent states of prior sensitization to non-HLA antigens. Non-HLA antigens are relatively weak when initially encountered and are therefore suppressible by conventional immunosuppressive therapy. Once priming has occurred, however, secondary responses are much more refractory to treatment. ABO incompatibilities are hazardous because of the presence of natural anti-A and anti-B antibodies in recipients and the normal expression of A and B blood group substances on endothelium, resulting in immediate vascular injury.

Living Donors When first-degree relatives are donors, graft survival rates at 1 year are slightly greater than those for cadaver grafts, with the exception of HLA-identical donors where 1-year results are approximately 95%. After the first year, the long-term survival rates as defined by the $t_{1/2}$still favor the partially matched (one HLA haplotype) family donor over a randomly selected cadaver donor (Table 272-1). In addition, living donors provide the advantage of immediate availability. Waiting lists for cadaveric kidneys have grown faster than the available organ supply, to the point where most new patients with end-stage renal disease wait for more than 4 years. In response to this increasing

disparity between cadaver donor supply and patient demand, living unrelated volunteers, usually spouses or close friends, are being accepted as donors in increasing numbers. It is illegal in the United States to purchase organs for transplantation. The results of transplantation using living unrelated donors have been most satisfactory, with initial and long-term survival rates the same as for partial HLA-matched family donors and better than for partially matched cadaveric donors (Table 272-1).

Concern has been expressed regarding the potential risk to a volunteer kidney donor of premature renal failure after several years of increased blood flow and hyperfiltration per nephron in the remaining kidney. There are a few reports of the development of hypertension, proteinuria, and even lesions of focal segmental sclerosis in donors under long-term follow-up. Difficulties in donors followed for 20 or more years are unusual, however, and it may be that having a single kidney becomes significant only when another condition, such as hypertension, is superimposed. It is also desirable to consider the risk of development of type 1 diabetes mellitus in a family member who is a potential donor to a diabetic renal failure patient. Anti-insulin and anti-islet antibodies should be measured, and glucose tolerance tests should be performed in such donors to rule out a prediabetic state.

HLA Matching and Cadaveric Donors The question of whether matching of HLA antigens in unrelated donor-recipient pairs would approximate the high initial success rates and slow rates of subsequent graft loss with HLA-identical sib pairs could not be answered until the late 1980s when reliable class II histocompatibility (DR) typing became widely available. Now that pooled data on tens of thousands of cadaveric renal transplants from all over the world are available, the HLA-matching effect can be clearly seen, especially in the long-term $t_{1/2}$half-life survival figures. It is shown inTable 272-1 that there is an overall beneficial effect of HLA matching in first cadaveric grafts. When compared with HLA-identical transplants, in which the 1-year graft survival rate is 95% and the subsequent half-life is 25 years, one-HLA-haplotype-matched family donor transplants have 1-year survival rates of 85% with a 12-year half-life (Table 272-1). With increasing numbers of mismatches for cadaveric donors, the half-life decreases from 20 to 7.7 years. The survival rates at the 10-year mark are projected to range from 65 (zero mismatches) to 34% (six mismatches). Many centers now report 1-year graft survival rates in the 85 to 90% range for all renal transplants (Fig. 272-1), possibly the result of heavy initial immunosuppression, but the subsequent half-lives are similar to those above. There is controversy regarding the value of cadaveric organ-sharing rules that are based entirely upon the numbers of HLA mismatches. Avoidance of mismatching for six antigens Table 272-1) is a top priority in the United States, however, and 20% of kidneys are transplanted on this basis.Table 272-1 also shows the interaction of HLA matching and graft ischemia on results; namely, kidneys from HLA-incompatible spousal donors do better than those from similarly mismatched cadaver donors, suggesting that the additional ischemic injury of organ storage is important. When such a cadaveric donor is HLA-compatible, however, ischemia and storage do not impede the matching benefit.

Presensitization A positive cross match of recipient serum with donor T lymphocytes representing anti-HLA class I is usually predictive of an acute vasculitic event termed *hyperacute rejection*. Patients with anti-HLA antibodies can be safely transplanted if careful cross matching of donor blood lymphocytes with recipient serum is performed.

Patients sustained by dialysis often show fluctuating antibody titers and specificity patterns. At the time of assignment of a cadaveric kidney, cross matches are performed with at least a current serum. Previously analyzed antibody specificities and additional cross matches are performed accordingly. Techniques for cross matching are not universally standardized; however, at least two techniques are employed in most laboratories. The minimal purpose for the cross match is avoidance of hyperacute rejection mediated by recipient antibodies to donor HLA class I antigens. Sensitive tests, such as the use of flow cytometry, can be useful for avoidance of accelerated, and often untreatable, early graft rejection in patients receiving second or third transplants. Donor T lymphocytes, which express only class I antigens, are used as targets for detection of anti-class I (HLA-A and -B) antibodies. Anti-class II (HLA-DR) antibodies do not contraindicate transplantation, unless present in high titer. B lymphocytes expressing both class I and class II antigens are used in these assays. Non-HLA antigens restricted in expression to endothelium and sometimes monocytes have been described, but clinical relevance is not well established.

Blood Transfusions Exposure to leukocyte HLA antigens during transfusions is a major cause of sensitization that limits transplantation access and increases the risk of early graft rejection. In the 1970s, attempts to avoid all blood exposure in dialysed patients paradoxically increased the risk of graft rejection. The beneficial "transfusion effect" was never fully explained, and it almost disappeared in the 1980s as overall management of patients improved with the use of cyclosporine and more effective means of rejection treatment. Currently, with the use of erythropoietin the need for transfusion is much reduced. It has been noted, however, that nontransfused patients do have more rejection activity.

IMMUNOLOGY OF REJECTION

Knowledge of the immunology of tissue transplantation stems largely from animal experimentation. However, enough evidence has accumulated in humans to indicate that the mechanisms are not qualitatively different from those found in other areas of immunology (Chap. 305). Early rejection is associated with activation of T lymphocytes having direct specificity against donor antigens. These may be cytotoxic cells (CD8+ or CD4+) or cells that mediate delayed hypersensitivity (CD4+); however, significant numbers of B lymphocytes, natural killer cells, and macrophages appear in the early infiltrate, and cells capable of mediating antibody-dependent cell-mediated cytotoxicity are also present. Many of the B lymphocytes produce immunoglobulins. The spectrum of cellular and humoral response and graft injury is quite varied, depending on specific genetic differences between donor and recipient and states of presensitization. The greater the degree of presensitization, the more likely it is that one will find antibody-mediated vascular lesions. All the processes shown inFig. 272-2 are possible, but their relative contribution varies from case to case. Monitoring of peripheral blood lymphocyte subsets utilizing monoclonal antibodies to functionally related surface molecules, such as CD4 (T helper cells) and CD8 (T cytotoxic cells), has been related to the degree of rejection activity in some surveys. Since the principal role of the CD4 molecule is to promote interaction of T cells with class II HLA molecules on antigen-presenting cells and similarly CD8 interacts with class I HLA (Chap. 305), it is not surprising that both types of T cells are usually present. Finally, the cytokine mediators of the cellular immune response [interleukin (IL) 1 to IL-4, IL-6, IL-10, IL-12,

tumor necrosis factor (TNF), and interferon g] are involved in the control and expression of the alloimmune rejection response. For example, T cell production of interferon g causes increased expression of HLA antigens on endothelial cells. In normal immunobiology this effect may be to promote more efficient presentation of foreign antigen, while in transplantation it enhances the immunogenicity of the vascularized transplant. Also, IL-2, the major growth factor for expansion of effector T cells, is the product of a major subset of CD4 cells (Th1), while other CD4 cells (Th2) produce B cell growth factors, such as IL-4.

The failure of transplanted kidneys after several years of adequate function is said to be due to "chronic rejection." In such kidneys, the development of nephrosclerosis, with proliferation of the vascular intima of renal vessels, and intimal fibrosis, with marked decrease in the lumen of the vessels, takes place (Fig. 272-3). The result is renal ischemia, hypertension, tubular atrophy, interstitial fibrosis, and glomerular atrophy with eventual renal failure. It is not established, however, whether slow deterioration of graft function over years is due to the same mechanisms in all cases. In addition to the established influence of HLA incompatibility, the age, number of nephrons, and ischemic history of a donor kidney may contribute to ultimate progressive renal failure in transplanted patients.

IMMUNOSUPPRESSIVE TREATMENT

Immunosuppressive therapy, as presently available, generally suppresses all immune responses, including those to bacteria, fungi, and even malignant tumors. In the 1950s when clinical renal transplantation began, sublethal total-body irradiation was employed. We have now reached the point where sophisticated pharmacologic immunosuppression is available, but it still has the hazard of promoting infection and malignancy. In general, all clinically useful drugs are more selective to primary than to memory immune responses. Agents to suppress the immune response are discussed in the following paragraphs, and those currently in clinical use are listed inTable 272-2.

Drugs *Azathioprine*, an analogue of mercaptopurine, was for two decades the keystone to immunosuppressive therapy in humans. This agent can inhibit synthesis of DNA, RNA, or both. Because cell division and proliferation are a necessary part of the immune response to antigenic stimulation, suppression by this agent may be mediated by the inhibition of mitosis of immunologically competent lymphoid cells, interfering with synthesis of DNA. Alternatively, immunosuppression may be brought about by blocking the synthesis of RNA (possibly messenger RNA), inhibiting processing of antigens prior to lymphocyte stimulation. Therapy with azathioprine in doses of 1.5 to 2.0 mg/kg per day is generally added to cyclosporine as a means of decreasing the requirements for the latter. Because azathioprine is rapidly metabolized by the liver, its dosage need not be varied directly in relation to renal function, even though renal failure results in retention of the metabolites of azathioprine. Reduction in dosage is required because of leukopenia and occasionally thrombocytopenia. Excessive amounts of azathioprine may also cause jaundice, anemia, and alopecia. If it is essential to administer allopurinol concurrently, the azathioprine dose must be reduced, since inhibition of xanthine oxidase delays degradation. This combination is best avoided.

Mycophenolate mofetil is now used in place of azathioprine in many centers. It has a

similar mode of action and a mild degree of gastrointestinal toxicity but produces minimal bone marrow suppression. Its advantage is its increased potency in preventing or reversing rejection.

Glucocorticoids are important adjuncts to immunosuppressive therapy. Of all the agents employed, prednisone has effects that are easiest to assess, and in large doses it is usually effective for the reversal of rejection. In general, 200 to 300 mg prednisone is given immediately prior to or at the time of transplantation, and the dosage is reduced to 30 mg within a week. The side effects of the glucocorticoids, particularly impairment of wound healing and predisposition to infection, make it desirable to taper the dose as rapidly as possible in the immediate postoperative period. Customarily, methylprednisolone, 0.5 to 1.0 g intravenously, is administered immediately upon diagnosis of beginning rejection and continued once daily for 3 days. When the drug is effective, the results are usually apparent within 96 h. Such "pulse" doses are not effective in chronic rejection. Most patients whose renal function is stable after 6 months or a year do not require large doses of prednisone; maintenance doses of 10 to 15 mg/d are the rule. Many patients tolerate an alternate-day course of steroids without an increased risk of rejection.

A major effect of steroids is on the monocyte-macrophage system, preventing the release ofIL-6 and IL-1. Lymphopenia after large doses of glucocorticoids is primarily due to sequestration of recirculating blood lymphocytes to lymphoid tissue.

Cyclosporine is a fungal peptide with potent immunosuppressive activity. It acts on the calcineurin pathway to block transcription of mRNA for IL-2 and other proinflammatory cytokines, thereby inhibiting T cell proliferation. Although it works alone, cyclosporine is more effective in conjunction with glucocorticoids. Since cyclosporine blocks production of IL-2 by T cells, its combination with steroids is expected to produce a double block in the macrophage ®IL-6/IL-1 ® T cell ®IL-2 sequence. As noted, clinical results with tens of thousands of renal transplants have been impressive. Of its toxic effects (nephrotoxicity, hepatoxicity, hirsutism, tremor, gingival hyperplasia, diabetes), only nephrotoxicity presents a serious management problem and is further discussed below.

Tacrolimus (FK-506) is a fungal macrolide that has the same mode of action, and a similar side effect profile, as cyclosporine. It does not produce hirsutism or gingival hyperplasia, however. De novo induction of diabetes mellitus is more common with tacrolimus. The drug was first used in liver transplantation, and may substitute for cyclosporine entirely, or be tried as an alternative in renal patients whose rejections are poorly controlled by cyclosporine.

Sirolimus (previously called rapamycin) is another fungal macrolide but has a different mode of action: namely, it inhibits T cell growth factor pathways, preventing the response toIL-2 and other cytokines. It shows some promise in clinical trials in combination with cyclosporine.

Antibodies to Lymphocytes When serum from animals made immune to host lymphocytes is injected into the recipient, a marked suppression of cellular immunity to the tissue graft results. The action on cell-mediated immunity is greater than on humoral immunity. A globulin fraction of serum [antilymphocyte globulin (ALG)] is the agent

generally employed. For use in humans, peripheral human lymphocytes, thymocytes, or lymphocytes from spleens or thoracic duct fistulas have been injected into horses, rabbits, or goats to produce antilymphocyte serum, from which the globulin fraction is then separated. Monoclonal antibodies against defined lymphocyte subsets offer a more precise and standardized form of therapy. OKT3 is directed to the CD3 molecules that form a portion of the T cell antigen-receptor complex; hence CD3 is expressed on all mature T cells. CD4 or CD8 molecules also form part of the fully activated cluster of molecules, and monoclonal antibodies to these offer the potential for more selective targeting of T cell subsets. Another approach to more selective therapy is to target the 55-kDa alpha chain of theIL-2 receptor, expressed only on T cells that have been recently activated. The problem with such mouse antibodies is the potential for developing human antimouse antibodies (HAMA), an event that limits the effective period of use. Genetically engineered monoclonal antibodies can solve this problem. Two such antibodies to the IL-2 receptor, in which either a chimeric protein has been made between mouse Fab with human Fc (basiliximab) or "humanized" by splicing the combining sites of the mouse into a molecule that is 90% human IgG (daclizumab), have been approved for use, after clinical evidence of reduction of rejection episodes. Their precise clinical role is under study.

CLINICAL COURSE AND MANAGEMENT OF THE RECIPIENT

Adequate hemodialysis should be performed within 48 h of surgery, and care should be taken that the serum potassium level is not markedly elevated so that intraoperative cardiac arrhythmias can be averted. The diuresis that commonly occurs postoperatively must be carefully monitored; in some instances it may be massive, reflecting the inability of ischemic tubules to regulate sodium and water excretion; with large diureses, massive potassium losses may occur. Most chronically uremic patients have some excess of extracellular fluid, and it is useful to maintain an expanded fluid volume in the immediate postoperative period. Acute tubular necrosis (ATN) may cause immediate oliguria or may follow an initial short period of graft function. ATN is most likely when cadaveric donors have been hypotensive or if the interval between cessation of blood flow and organ harvest (warm ischemic time) is more than a few minutes. Recovery usually occurs within 3 weeks, although periods as long as 6 weeks have been reported. Superimposition of rejection on ATN is common, and the differential diagnosis may be difficult without a graft biopsy. Cyclosporine therapy prolongs ATN, and some patients do not diurese until the dose is drastically reduced. Many centers avoid starting cyclosporine for the first several days, usingALG or a monoclonal antibody along with mycophenolate mofetil and prednisone until renal function is established.

The Rejection Episode Early diagnosis of rejection allows prompt institution of therapy to preserve renal function and prevent irreversible damage. Clinical evidence of rejection is rarely characterized by fever, swelling, and tenderness over the allograft. Rejection may present only with a rise in serum creatinine, with or without a reduction in urine volume. The focus should be on ruling out other causes of functional deterioration.

Arteriography and radioactive iodohippurate sodium renograms of the transplanted kidney may be useful in ascertaining changes in the renal vasculature and in renal blood flow, even in the absence of urinary flow. Thrombosis of the renal vein occurs rarely; it may be reversible if caused by technical factors and intervention is prompt. Diagnostic

ultrasound is the procedure of choice to rule out urinary obstruction or to confirm the presence of perirenal collections of urine, blood, or lymph. When renal function has been good initially, a rise in the serum creatinine level is the most sensitive and reliable indicator of possible rejection and may be the only sign.

Calcineurin inhibitors (cyclosporine or tacrolimus) may cause deterioration in renal function in a manner similar to a rejection episode. In fact, rejection processes tend to be more indolent with these inhibitors, and the only way to make a diagnosis may be by renal biopsy. Calcineurin inhibitors have an afferent arteriolar constrictor effect on the kidney and may produce permanent vascular and interstitial injury after sustained high-dose therapy. Addition of angiotensin-converting enzyme (ACE) inhibitors or nonsteroidal anti-inflammatory drugs are likely to raise serum creatinine levels. The former are generally safe to use after the early months, while the latter are best avoided in all renal transplant patients. There is no universally accepted lesion(s) that makes a diagnosis of calcineurin inhibitor toxicity, although interstitial fibrosis, isometric tubular vacuolization, and thickening of arteriolar walls have been noted by some. Basically, if the biopsy does not reveal moderate and active cellular rejection activity, the serum creatinine will most likely respond to a reduction in dose. Blood levels of drug can be useful if very high or very low but do not correlate precisely with renal function, although serial changes in a patient can be useful. If rejection activity is present in the biopsy, appropriate therapy is indicated. The first rejection episode is usually treated with intravenous administration of methylprednisolone, 500 to 1000 mg daily for 3 days. Failure to respond is indication for antibody therapy, usually with OKT3.

OKT3 monoclonal antibody, given intravenously for 10 to 14 days, is effective in more than 90% of first rejections, and less so if methylprednisolone pulses have failed and in cases of severe recurrent rejection activity. A major problem with OKT3 is that severe systemic reactions may be produced during the first day or two of therapy. Chills, fever, hypotension, and headache are the direct result of the antibody effects on the targeted T cells, most likely related to the known potential of OKT3 to activate T cells nonspecifically with release of cytokines, especially TNF-a. If the antibody is administered to overhydrated oliguric patients, pulmonary edema may be induced. These reactions are not characteristic of other monoclonal antibodies, such as those to theIL-2 receptor. Recurrent or rebound rejection activity may require additional therapy. In such circumstances, methylprednisolone may be effective even though it failed initially. Second courses of OKT3 may be given in spite ofHAMAgenerated in response to the first course if the titers are low and the human antibodies are not directed to the combining-site region (idiotype) of the OKT3.

Management Problems The usual clinical manifestations of infection in the posttransplant period are blunted by immunosuppressive therapy. The major toxic effect of azathioprine is bone marrow suppression, which is less likely with mycophenolate mofetil, while calcineurin inhibitors have no marrow effects. All drugs predispose to unusual opportunistic infections, however. The signs and symptoms of infection may be masked and distorted, and fever without obvious cause is common. Only after days or weeks it may become apparent that it has a viral or fungal origin. Bacterial infections are most common during the first month after transplantation. The importance of blood cultures in such patients cannot be overemphasized, because systemic infection without obvious foci is frequent, although wound infections with or without urinary fistulas are

most common. Particularly ominous are rapidly occurring pulmonary lesions, which may result in death within 5 days of onset. When these become apparent, immunosuppressive agents should be discontinued, except for maintenance doses of prednisone. Aggressive diagnostic procedures, including transbronchial and open lung biopsy, are frequently indicated. In the case of *Pneumocystis carinii* (Chap. 209) infection, trimethoprim-sulfamethoxazole is the treatment of choice; amphotericin B has been used effectively in systemic fungal infections. Prophylaxis against *P. carinii* with daily, or alternate day, low-dose trimethoprim-sulfamethoxazole is very effective. Involvement of the oropharynx with *Candida* (Chap. 205) may be treated with local nystatin. Tissue-invasive fungal infections require treatment with systemic agents such as fluconazole. Small doses (a total of 300 mg) of amphotericin given over a period of 2 weeks may be effective in fungal infections refractory to fluconazole. Macrolide antibiotics, especially ketoconazole and erythromycin, and some calcium channel blockers (diltiazem, verapamil) compete with calcineurin inhibitors for P450 catabolism and cause elevated levels of these immunosuppressive drugs. Analeptics, such as phenytoin and carbamazepine, will increase catabolism to result in low levels. *Aspergillus* (Chap. 206), *Nocardia* (Chap. 165), and cytomegalovirus (CMV) (Chap. 185) infections also occur.

CMVis a common and dangerous infection in transplant recipients. It does not generally appear until the end of the first posttransplant month. Active CMV infection is sometimes associated, or occasionally confused, with rejection episodes. Patients at highest risk for severe CMV disease are those without anti-CMV antibodies who receive a graft from a CMV antibody-positive donor (15% mortality). Serial intravenous administration of high-titer CMV immune globulin is effective in reducing this risk. Prophylactic use of ganciclovir is an effective alternative. Early diagnosis in a febrile patient can be made by detecting CMV antigens in the blood. A rise in IgM antibodies to CMV is also diagnostic. Culture of CMV from blood may be less sensitive. Tissue invasion of CMV is common in the gastrointestinal tract and lungs. CMV retinopathy occurs late in the course, if untreated. Treatment of active CMV disease with ganciclovir is always indicated. Many patients immune to CMV can activate the virus after heavy immunosuppression, such as with OKT3. Concurrent treatment with ganciclovir during OKT3 administration appears to be effective for prophylaxis of CMV activation. The complications of glucocorticoid therapy are well known and include gastrointestinal bleeding, impairment of wound healing, osteoporosis, diabetes mellitus, cataract formation, and hemorrhagic pancreatitis. The treatment of unexplained jaundice in transplant patients should include cessation or reduction of immunosuppressive drugs if hepatitis or drug toxicity is suspected. It is surprising that cessation of azathioprine or calcineurin inhibitor therapy in such circumstances often does not result in rejection of a graft, at least for several weeks. Acyclovir is effective in therapy of herpes simplex virus infections.

Antiplatelet agents and anticoagulants, although effective in theory, have not been successful in the prevention of the "chronic rejection" vascular lesions. Persistent elevation of serum creatinine levels above 220 umol/L (2.5 mg/dL) in patients on calcineurin inhibitor is an indication for dose reduction, particularly if calcineurin inhibitor blood levels are elevated. The risk of long-term cumulative toxicity to the kidney now seems to be low. In general, minimal or no rejection during the first 6 months after transplantation is a predictor of safety in reducing immunosuppression therapy over subsequent months to years, but chronic progressive vasculopathy may still occur.

Despite the potential teratogenic effects of immunosuppressive agents, both women and men have become parents after transplantation. The incidence of congenital abnormalities in the offspring is not increased.

Glomerular Lesions Glomerular lesions occur in 10 to 15% of allografts, even when the original disease was accidental removal of a solitary kidney. The pathogenesis is related to a chronic rejection process. In some cases the lesions resemble those of the original glomerular disease. In most instances, the recurrence of the original renal lesions represents no threat to the immediate prognosis, and a primary diagnosis of glomerulonephritis is rarely a contraindication to transplantation. Focal segmental glomerulosclerosis may recur up to 30% of the time, with one-third of these patients losing graft function. Hemolytic uremic syndrome also has a high recurrence rate.

Malignancy The incidence of tumors in patients on immunosuppressive therapy is 5 to 6%, or approximately 100 times greater than that in the general population of the same age range. The most common lesions are cancer of the skin and lips and carcinoma in situ of the cervix, as well as lymphomas, such as non-Hodgkin's lymphomas. The risks are increased in proportion to the total immunosuppressive load administered and time elapsed since transplantation. Surveillance for skin and cervical cancers is necessary.

Other Complications *Hypercalcemia* after transplantation may indicate failure of hyperplastic parathyroid glands to regress. Aseptic necrosis of the head of the femur is probably due to preexisting hyperparathyroidism, with aggravation by glucocorticoid treatment. With improved management of calcium and phosphorus metabolism during chronic dialysis, the incidence of parathyroid-related complications has fallen dramatically. Persistent hyperparathyroid activity may require subtotal parathyroidectomy.

Hypertension may be caused by (1) native kidneys; (2) rejection activity in the transplant; (3) renal artery stenosis, if an end-to-end anastomosis was constructed with an iliac artery branch; and (4) renal calcineurin inhibitor toxicity. The latter may improve with reduction in dose. WhereasACEinhibitors may be useful, calcium channel blockers are more frequently used initially. Amelioration of hypertension to the 120-130/70-80 mmHg range should be the goal in all patients.

Chronic hepatitis, particularly when due to hepatitis B virus, can be a progressive, fatal disease over a decade or so. Patients who are persistently hepatitis B surface antigen-positive are at higher risk, according to some studies, but the presence of hepatitis C virus is also a concern when one embarks on a course of immunosuppression in a transplant recipient.

Both chronic dialysis and renal transplant patients have a higher incidence of death from myocardial infarction and stroke than in the population at large, and this is particularly true in diabetic patients. Contributing factors are the use of glucocorticoids, hypertension, and hypertriglyceridemia. Increased low-density lipoprotein cholesterol and depressed high-density lipoprotein cholesterol concentrations may be exaggerated after transplantation and require treatment. Recipients of renal transplants have a high prevalence of coronary artery and peripheral vascular diseases. The percentage of

deaths from these causes has been slowly rising as the numbers of transplanted diabetic patients and the average age of all recipients increase. More than 50% of renal recipient mortality is attributable to cardiovascular disease. In addition to strict control of blood pressure and blood lipid levels, close monitoring of patients for indications of further medical or surgical intervention is an important part of management.

(Bibliography omitted in Palm version)

Back to Table of Contents

273. PATHOGENESIS OF GLOMERULAR INJURY - *Hugh R. Brady*, *Barry M. Brenner*

The glomerulus is a modified capillary network that delivers an ultrafiltrate of plasma to Bowman's space, the most proximal portion of the renal tubule. Approximately 1.6 million glomeruli are present in two mature kidneys (range 0.5 to 2.4 million) and collectively they produce 120 to 180 L of ultrafiltrate daily. Glomerular filtration rate (GFR) is dependent on glomerular blood flow, ultrafiltration pressure, and surface area. These parameters are tightly regulated through changes in afferent and efferent arteriolar tone (for blood flow and ultrafiltration pressure) and mesangial cell contractility (for filtration surface area). Arteriolar tone and mesangial cell contractility are, in turn, modulated by neurohumoral factors, local myenteric reflexes, and endothelium-derived vasoactive substances, such as nitric oxide, prostacyclin, and endothelins. In health, glomerular endothelium is also antithrombotic and antiadhesive for leukocytes and platelets, thereby preventing inappropriate vascular thrombosis and inflammation during the filtration process. Filtration of most plasma proteins and all blood cells is normally prevented as a consequence of the physiochemical and electrostatic charge characteristics of the glomerular filtration barrier, the latter being composed of fenestrated glomerular endothelium, basement membrane, and the foot processes and slit diaphragms of visceral epithelial cells (podocytes). Parietal epithelium facilitates glomerular filtration by maintaining the integrity of Bowman's space. In keeping with the physiologic functions of the glomerulus outlined above, virtually all glomerular injury results in impairment of glomerular filtration and/or the inappropriate appearance of plasma proteins and blood cells in the urine.

CLINICOPATHOLOGIC CORRELATES IN GLOMERULAR DISEASE

The major glomerulopathies are described in Chap. 274, and major morphologic patterns of glomerular disease and their clinical features are summarized in Table 273-1. These clinicopathologic entities can be induced by a variety of different pathogenetic mechanisms. Thus, prompt diagnosis, optimal management, and accurate prognostication is a multistep process that requires (1) recognition of the presenting clinical syndrome, (2) delineation of the underlying morphologic pattern of glomerular injury, and (3) elucidation of the specific renal-limited or systemic disease that triggered glomerular dysfunction.

NOMENCLATURE

The terms *glomerulonephritis* and *glomerulopathy* are usually used interchangeably to denote glomerular injury, although some authorities reserve the former term for injury with evidence of inflammation such as leukocyte infiltration, antibody deposition, and/or complement activation. Glomerular diseases are classified as *primary* when the pathology is confined to the kidney and any systemic features are a direct consequence of glomerular dysfunction (e.g., pulmonary edema, hypertension, the uremic syndrome). Usually, but not always, the term primary is synonymous with *idiopathic*. Glomerular diseases are classified as *secondary* when part of a multisystem disorder. In general, *acute* refers to glomerular injury occurring over days or weeks, *subacute* or *rapidly progressive* over weeks or a few months, and *chronic* over many months or years. Lesions are classified as *focal* or *diffuse* when they involve the minority (<50%) or

majority (³50%) of glomeruli, respectively. Lesions are termed *segmental* or *global* when they involve part of or almost all of the glomerular tuft, respectively. *Proliferative* is used to describe an increase in glomerular cell number, which can be due to infiltration by leukocytes or proliferation of resident glomerular cells. Proliferation of resident glomerular cells is classified as *intracapillary* or *endocapillary* when referring to endothelial or mesangial cells and *extracapillary* when referring to cells in Bowman's space. A *crescent* is a half-moon-shaped collection of cells in Bowman's space, usually composed of proliferating parietal epithelial cells and infiltrating monocytes. Because crescentic glomerulonephritis is often associated with renal failure that progresses rapidly over week to months, the clinical term *rapidly progressive glomerulonephritis* and pathologic term *crescentic glomerulonephritis* are often used interchangeably. The description *membranous* is applied to glomerulonephritis dominated by expansion of the glomerular basement membrane (GBM) by immune deposits. *Sclerosis* refers to an increase in the amount of homogeneous nonfibrillar extracellular material of the same ultrastructural appearance and chemical composition as GBM and mesangial matrix. This process is distinct from *fibrosis*, which involves deposition of collagens type I and III and is more commonly a consequence of healing of crescents or tubulointerstitial inflammation.

MAJOR CLINICOPATHOLOGIC ENTITIES

Most glomerulopathies are still classified and named according to their morphologic features (Table 273-1). The major *inflammatory glomerulopathies* are focal proliferative glomerulonephritis (termed *mesangial proliferative* if the proliferating cells are predominantly mesangial cells), diffuse proliferative glomerulonephritis, and crescentic glomerulonephritis. These diseases typically present with a *nephritic-type* "active" urine sediment characterized by the presence of red blood cells, red blood cell casts, leukocytes, and *subnephrotic* proteinuria of <3 g/24 h. The severity of renal insufficiency varies in proportion to the degree of proliferation and necrosis.

The major morphologic patterns affecting the glomerular filtration barrier for proteins, namely theGBM and visceral epithelial cells, are membranous glomerulopathy, minimal change disease, and focal and segmental glomerulosclerosis. These entities typically present with *nephrotic-range* proteinuria of>3 g/24 h and the presence of few red blood cells, leukocytes, or cellular casts. As a consequence of the heavy proteinuria, nephrotic syndrome is associated with hypoalbuminemia, edema, hyperlipidemia, and lipiduria. Membranoproliferative glomerulonephritis, as the name suggests, is a hybrid lesion that presents with a combination of nephritic and nephrotic features.

The *glomerular deposition diseases* are a group of disorders characterized by prominent extravascular deposition of a paraprotein or fibrillar material. These diseases can also trigger nephritic-type and nephrotic-type responses (or a combination of both) and thus show marked clinical and morphologic overlap with the entities described above.

The *thrombotic microangiopathies* are a family of diseases in which the pathologic presentation is dominated by thrombi within the renal microvasculature, often leading to renal insufficiency.

MAJOR DETERMINANTS OF GLOMERULAR INJURY

Important determinants of the severity of glomerular injury include (1) the nature of the primary insult and the secondary mediator systems that it invokes, (2) the site of injury within the glomerulus; and (3) the speed of onset, extent, and intensity of disease.

PRIMARY INSULT

Glomeruli are susceptible to a variety of inflammatory, metabolic, hemodynamic, toxic, and infectious insults (Table 273-2). Most human glomerular disease is triggered by immune attack, diabetes mellitus, or hypertension. Diverse insults can induce similar clinicopathologic presentations, suggesting marked overlap among downstream molecular and cellular responses. For example, infections (e.g., streptococcal pharyngitis, bacterial endocarditis) and vasculitides (e.g., Henoch-Schonlein purpura, microscopic polyarteritis) can each trigger acute proliferative glomerulonephritis with the nephritic syndrome. Similarly, metabolic (e.g., diabetes mellitus) and deposition diseases (e.g., amyloid) can each induce glomerulosclerosis with nephrotic syndrome. An important corollary is that pharmacologic agents that inhibit common secondary mediator systems may prove effective in treating glomerular diseases of diverse etiologies (see below).

SITE OF INJURY

The consequences of injury at different sites within the glomerulus can be predicted from the physiologic functions of the cells within the local milieu (Table 273-3). The major sequelae of injury to the *endothelium* and *subendothelial aspect of theGBM*are (1) recruitment of leukocytes leading to inflammatory glomerulonephritis, (2) perturbed hemostasis leading to thrombotic microangiopathy, and (3) vasoconstriction and mesangial cell contraction leading to acute renal failure. It is usual for one of these phenotypes to dominate the presentation of specific diseases. *Mesangial* injury is usually immunologic in origin and, being more localized, induces less dramatic impairment of glomerular filtration. Patients typically present with asymptomatic abnormalities of the urinary sediment and mild renal insufficiency. Proteinuria dominates the clinical presentation of injury to the *subepithelial aspect of the GBM* and *visceral epithelial cells*. As with mesangial injury,GFR is often only mildly compromised in this setting. The classic pathologic manifestation of *parietal epithelial cell* injury is crescent formation. Crescents can be the dominant morphologic presentation of glomerular disease or complicate proliferative or membranous lesions.

SPEED OF ONSET, INTENSITY, AND EXTENT OF INJURY

To illustrate the importance of the speed of onset, extent, and intensity of glomerular injury, it is instructive to compare two forms of immune complex glomerulonephritis, namely, acute postinfectious glomerulonephritis and IgA nephropathy. Postinfectious glomerulonephritis is characterized by rapid and extensive formation of immune complexes throughout the glomerular capillary wall, which often provokes acute renal failure with the classic hallmarks of acute inflammation: complement activation, leukocyte recruitment, lysosomal enzyme release, free radical generation, and perturbation of vascular tone and permeability. In contrast, IgA nephropathy is characterized by slow, but sustained, formation of immune complexes, largely confined

to the mesangium; less dramatic activation of complement and other secondary mediator systems; and either stability ofGFR or progressive renal insufficiency over 10 to 20 years.

IMMUNOLOGIC GLOMERULAR INJURY

Immune-mediated glomerulonephritis (Chaps. 274 and275) accounts for a large fraction of acquired renal disease. The majority of cases are associated with the deposition of antibodies, often autoantibodies, within the glomerular tuft, indicating dysregulation of humoral immunity. Cellular immune mechanisms also contribute to the pathogenesis of antibody-mediated glomerulonephritis by modulating antibody production and through antibody-dependent cell cytotoxicity (see below). In addition, cellular immune mechanisms probably play a primary role in the pathophysiology of "pauci-immune" glomerulonephritides, notable for robust glomerular inflammation in the absence of immunoglobulin deposition.

ANTIBODY-MEDIATED INJURY (Fig. 273-1)

Most antibody-mediated glomerulonephritis in humans is initiated by reactivity of circulating antibodies with auto- or "planted" antigens within the glomerulus. The major mechanisms of antibody deposition within the glomerulus are (1) reactivity of circulating autoantibodies with intrinsic autoantigens that are components of normal glomerular parenchyma, (2) in situ formation of immune complexes through interaction of circulating antibodies with extrinsic antigens that have been planted within the glomerulus, and (3) intraglomerular trapping of immune complexes that have formed in the systemic circulation. Autoantibodies against neutrophil cytoplasmic antigens in the circulation may represent an additional mechanism of antibody-mediated glomerular injury in patients without discernible immune complexes in the glomerular parenchyma (see below).

Generation of Nephritogenic Antibodies Exposure of the host to a foreign antigen (e.g., a prodromal infection) has been implicated as the trigger for the generation of nephritogenic autoantibodies in several forms of glomerulonephritis. Foreign antigens can provoke autoantibody formation through several mechanisms. First, a foreign antigen, whose structure resembles that of a host glomerular antigen, may stimulate the production of autoantibodies that cross-react with the intrinsic glomerular antigen ("molecular mimickry"). Second, the foreign antigen may trigger aberrant expression of major histocompatibility complex class II molecules on glomerular cells which present previously "invisible" autoantigens to T lymphocytes and thereby generate an autoimmune response. Third, the foreign antigen can trigger polyclonal activation of B lymphocytes, some of which generate nephritogenic antibodies. Alternatively, individuals may suffer a breakdown of immune tolerance through other mechanisms (e.g., genetically programmed). Autoreactive B cells are usually deleted in the thymus during development (clonal deletion) or rendered anergic in peripheral lymphoid tissue (clonal anergy). Similar tolerogenic mechanisms exist for deleting or anergizing autoreactive T helper cells that modulate immunoglobulin production by autoreactive B cells. Perturbation of either of these tolerogenic mechanisms could drive immunoglobulin production in some forms of autoimmune glomerulonephritis. Indeed, defective clonal deletion of autoreactive T cells has been demonstrated in experimental

lupus nephritis due to defective synthesis of Fas, a cell-surface receptor that modulates T cell deletion through apoptosis (programmed cell death) within the thymus.

Deposition of Nephritogenic Antibodies within the Glomerulus (Fig. 273-1)
Anti-GBMantibody disease (p. 1583) is the classic nephritis initiated by interaction of autoantibody with intrinsic glomerular antigen. Afflicted patients have a circulating antibody directed at a 28-kDa antigen (Goodpasture antigen) located in the noncollagenous NC1 domain of the a3 chain of type IV collagen. This type of collagen is preferentially expressed in glomerular and pulmonary alveolar basement membranes. Autoantibodies against mesangial cell antigens have been detected in the serum of patients with IgA nephropathy, the most common form of glomerulonephritis in humans; however, the pathogenicity of these autoantibodies has yet to be defined. Poststreptococcal glomerulonephritis and lupus nephritis are examples of glomerulonephritides that are probably initiated by interaction of circulating antibodies with planted antigens. Several streptococcal antigens have been isolated from immune deposits of kidneys with poststreptococcal glomerulonephritis, including nephritis strain-associated antigen and a cytoplasmic protein endostreptosin. In addition, patients with poststreptococcal glomerulonephritis can have circulating antibodies against laminin, type IV collagen, and heparan sulphate proteoglycans, suggesting that molecular mimickry may also contribute. Similar findings have been reported in experimental and human lupus nephritis. Here, circulating anti-DNA antibodies may potentially induce immune complex glomerulonephritis by reacting with DNA bound to GBM or with planted DNA-histone complexes (nucleosomes). However, it should be noted that patients with systemic lupus erythematosus have a variety of circulating autoantibodies, and the pathogenetic culprit(s) in lupus nephritis have yet to be identified definitively. Cryoglobulinemia, due to chronic hepatitis C infection, is an example of glomerulonephritis initiated by trapping of immune complexes. These patients have circulating and intraglomerular immune complexes composed of hepatitis C antigens, polyclonal antihepatitis C IgG, and a second antibody, usually a monoclonal IgM, directed against the IgG. In support of the pathogenicity of circulating cryoglobulins, their injection into laboratory mice induces glomerulonephritis with many of the hallmarks of human disease.

Site of Antibody Deposition The site of antibody deposition within the glomerulus is a critical determinant of the clinicopathologic presentation. Among the factors that determine the site of deposition are the avidity, affinity, and quantity of the antibody; the size, charge, and site of the antigen; the size of the immune complexes; the efficiency of the clearance mechanisms for immune complexes; and local hemodynamic factors. Relatively anionic antigens are repelled by theGBM, which is negatively charged, and tend to be trapped in the subendothelial cell space and mesangium. In contrast, relatively cationic antigens tend to permeate the GBM and deposit within the GBM or in the subepithelial space. Acute deposition of antibody in the subendothelial cell space or mesangium typically triggers a nephritic-type response characterized by rapid recruitment of leukocytes and platelets, probably because inflammatory mediators generated at these sites are strategically positioned to activate endothelial and hematogenous cells. Inflammation is more severe when antibody is deposited in the subendothelial space, as compared with mesangium, at least in part because the mesangium abuts only 25 to 33% of the capillary wall. Antibody deposition in the subepithelial cell space typically induces a nephrotic-type response characterized by

proteinuria without a pronounced inflammatory cell infiltrate, probably because the immune complexes are shielded from circulating inflammatory cells by the GBM and because the large fluid flux from blood to Bowman's space minimizes back-diffusion of inflammatory mediators towards the endothelium and vascular lumen.

Recruitment of Inflammatory Cells (Fig. 273-1) Leukocytes and platelets are important mediators of injury in most forms of acute and subacute glomerulonephritis. Immunoglobulin can provoke recruitment of leukocytes through several mechanisms. Many antibody subclasses activate the complement cascade, and complement proteins such as C3a, C5a, and C5b-9 (membrane attack complex) are potent stimuli for leukocyte recruitment, either through their direct effects on leukocytes (C3a, C5a) or by increasing endothelial cell adhesiveness for leukocytes (C5b-9). Complement-independent mechanisms also contribute. Leukocytes express Fc receptors that can directly engage the Fc portion of immunoglobulin. Resident glomerular macrophages, endothelial cells, and mesangial cells also express Fc receptors, engagement of which can trigger release of an array of inflammatory mediators and chemotactic cytokines (chemokines) that promote directed locomotion of leukocytes (chemotaxis), binding of leukocytes to inflamed endothelium through cell surface leukocyte adhesion molecules, and diapedesis of leukocytes to the extravascular space.

The mechanisms of platelet recruitment in glomerulonephritis are less well defined. Potential mechanisms include direct binding of platelet Fc receptors with immunoglobulin, and interactions of platelets with endothelium, trapped leukocytes, collagen, and other components of exposedGBM, and with products of the coagulation cascade such as fibrin.

Mediators of Glomerular Injury (Fig. 273-2) *Nephritic-type antibody-mediated glomerular injury* is a vivid example of host defense gone awry. In normal host defense, leukocytes engulf microorganisms into phagosomes, which then fuse with intracellular lysosomes. Microorganisms are destroyed within phagolysosomes through the actions of free radicals, proteolytic enzymes, and other toxic molecules. This process facilitates killing with relative protection and preservation of host tissue. When host defense is inappropriately activated in autoimmune diseases, the inciting antigens are often fixed to (planted antigens) or are a component of host tissue (autoantigen). As a result, phagocytosis is less efficient ("frustrated phagocytosis"), and there is release of toxic moieties such as oxidants and proteases into the parenchyma where they destroy host cells and matrix components. In addition, cytotoxic T lymphocytes and natural killer cells can damage resident glomerular cells by releasing toxic compounds, such as perforins, a process that is facilitated by binding of these cytotoxic cells to glomerular cells through HLA molecules, Fc portions of immunoglobulin (antibody-dependent cell cytotoxicity), and other immune recognition systems. Platelets promote nephritic injury by promoting leukocyte recruitment and intrarenal vasoconstriction and by triggering microthrombi formation. Cytokines, such as tumor necrosis factor a, interleukin 1b, and interferong, play a key role in the amplification and maintenance of glomerular inflammation by inducing de novo synthesis of leukocyte adhesion molecules, chemokines, and other inflammatory mediators.

Leukocytes play a lesser role in *nephrotic-type antibody-mediated glomerular injury*

(Chap. 274). Membranous glomerulopathy is the prototypic entity and is initiated by the formation of subepithelial immune complexes; these provoke production of "spikes" of new basement membrane that eventually encircle and incorporate the immune complexes into theGBM. The antigenic targets in human membranous glomerulopathy have not been determined but may be planted antigens or autoantigens shed from parietal epithelial cells. The frequent association with infections, malignancies, and drugs suggests involvement of planted antigens or molecular mimickry (see above); however, many cases may represent a true loss of tolerance against autoantigens. The membrane attack complex of complement (C5b-9) appears to be a major effector of injury to the glomerular filtration barrier in this setting.

CELL PROLIFERATION AND ACCUMULATION OF EXTRACELLULAR MATRIX

A hallmark of the nephritic-type proliferative glomerulopathies is an increase in glomerular cell number. Initially, this hypercellularity is due predominantly to infiltration of the glomerular tuft by leukocytes. Subsequently, resident glomerular cells proliferate in response to growth factors [e.g., epidermal growth factor, platelet-derived growth factor (PDGF), thrombospondin] released into the local inflammatory milieu. The proliferating cells are typically mesangial in mesangioproliferative glomerulonephritis and both endothelial and mesangial cells in diffuse proliferative glomerulonephritis. The visceral epithelial cell is, for the most part, a terminally differentiated cell that does not proliferate rapidly, even when injured.

Whereas acute antibody-mediated glomerulonephritis typically induces acute diffuse proliferative glomerulonephritis and acute renal failure over days to weeks (nephritic syndrome), subacute immune injury often induces the formation of glomerular crescents and renal failure over weeks-to-months (termed *rapidly progressive glomerulonephritis*). As discussed above, crescents are extracapillary proliferations of cells in Bowman's space, composed of infiltrating monocytes, proliferating parietal epithelial cells, and fibrin.

Sustained low level immune complex deposition over months to years can provoke a marked increase in basement membrane or mesangial matrix production. Mild to moderate accumulation of matrix usually manifests as proteinuria due to disruption of the glomerular filtration barrier; however, in its most severe form, matrix accumulation causes glomerulosclerosis and chronic renal insufficiency.

RESOLUTION, REPAIR, AND SCARRING

Glomerular inflammation can resolve with complete recovery of renal function or with a variable amount of scarring and chronic renal insufficiency. Acute poststreptococcal glomerulonephritis (p. 1582), for example, usually resolves spontaneously and fully in children, whereas adults are frequently left with residual renal impairment. The resolution process requires cessation of further antibody production and immune complex formation, removal of deposited and circulating immune complexes, inhibition of further recruitment of inflammatory cells, dissipation of the gradients of inflammatory mediators, restoration of normal endothelial adhesiveness and permeability, normalization of vascular tone, and clearance of infiltrating inflammatory cells and proliferating resident glomerular cells (Fig. 273-1).

Unfortunately, the resolution phase of most inflammatory glomerulopathies in adults terminates in some glomerular scarring. This is particularly true in patients with crescentic glomerulopathies who may be left with end-stage renal failure requiring dialysis or transplantation. Transforming growth factor (TGF)b, a cytokine, stimulates production of extracellular matrix by most glomerular cells, inhibits synthesis of tissue proteases that normally degrade matrix proteins, and is a potent stimulus for scar formation immediately following glomerular injury.

Moderate-to-severe glomerulonephritis is usually associated with a variable degree of tubulointerstitial inflammation and scarring in addition to glomerular injury. Indeed, the severity of tubulointerstitial injury usually correlates closely with long-term impairment of renal function. The pathogenesis of tubulointerstitial inflammation in this setting is unclear. Potential mechanisms include: (1) primary involvement of both the glomeruli and the tubulointerstitium in autoimmune disease; (2) induction of tubulointerstitial inflammation by mediators generated by diseased glomeruli which then diffuse into the tubulointerstitium via blood, tubular fluid, or the interstitial space; (3) injury to tubule epithelial cells by excessive filtered proteins ("protein overload" hypothesis); and (4) ischemia to areas of the tubulointerstitium downstream to areas of robust glomerular inflammation or severe glomerulosclerosis.

OTHER MECHANISMS OF ANTIBODY-MEDIATED INJURY

Several other autoantibodies have been implicated as mediators of renal injury in patients with glomerulonephritis.

Antineutrophil Cytoplasmic Antibodies (ANCA) Immunoglobulin is not detected in the glomerulus in approximately 40% of patients with rapidly progressive glomerulonephritis ("pauci-immune crescentic glomerulonephritis"). The majority of these patients have Wegener's granulomatosis, microscopic polyangiitis nodosa, or renal-limited crescentic glomerulonephritis and have autoantibodies against neutrophil cytoplasmic antigens in their circulation. When reactive with ethanol-fixed neutrophils isolated from healthy volunteers, ANCA stain results in either a cytoplasmic pattern (c-ANCA) or perinuclear pattern (p-ANCA). In the case of c-ANCA, the neutrophil antigen is usually proteinase-3, a constituent of neutrophil primary granules. In the case of p-ANCA, the antigen is usually myeloperoxidase, another granule constituent that migrates to the perinuclear area upon ethanol fixation. Whereas a greater number of patients with Wegener's granulomatosis have c-ANCA and a greater proportion of patients with renal-limited disease have p-ANCA, the morphologic features, response to treatment, and overall prognosis appear to be similar in patients with either c-ANCA or p-ANCA. ANCA stimulate cytokine-primed human neutrophils to generate reactive oxygen species and injure endothelium in vitro. These findings raise the possibility that ANCA may be pathogenetic in vivo in the presence of circulating cytokines, as may occur following a prodromal infection.

Antiendothelial Cell Antibodies Circulating antibodies against endothelial antigens have been reported in several inflammatory vasculitides and glomerulonephritides. Their titers tend to correlate with disease activity, and some activate endothelial cells and increase their adhesiveness for leukocytes, suggesting a pathogenetic role.

C3 Nephritic Factor Some patients with membranoproliferative glomerulonephritis (Chap. 273) have large deposits of electron-dense material within theGBM that does not stain for immunoglobulin (dense deposit disease; membranoproliferative glomerulonephritis type II). Intriguingly, most of these patients have a circulating IgG, termed the *C3 nephritic factor*, directed at C3bBb (C3 convertase) of the alternative pathway of complement.

CELL-MEDIATED INJURY

Although cell-mediated injury is, as yet, less well defined than antibody-mediated glomerular injury, T cells have also been implicated as independent mediators of glomerular injury and as modulators of the production of nephritogenic antibodies. T cells may be particularly important as initiators of injury in pauci-immune glomerulonephritis. T cells interact, through their cell-surface T cell receptor/CD3 complex, with antigens presented in the groove of major histocompatibility complex molecules of resident glomerular endothelial, mesangial, and epithelial cells, a process that is facilitated by cell-cell adhesion and costimulatory molecules. Cytokines and other mediators released by activated T cells are potent stimuli for further leukocyte recruitment, cytotoxicity, and fibrogenesis. CD4 T lymphocytes are important recruiters of macrophages and trigger clonal expansion of autoreactive B cells; they also promote glomerular cell injury by CD8 cytotoxic T lymphocytes and natural killer cells and through antibody-dependent cell cytotoxicity. Soluble factors derived from T cells have also been implicated in the pathogenesis of proteinuria in minimal change disease and primary focal segmental glomerulosclerosis. The identity and molecular characterization of these nonimmunoglobulin circulating permeability factors remain to be determined.

NONIMMUNOLOGIC GLOMERULAR INJURY

METABOLIC

Diabetic Nephropathy (See also Chaps. 275 and333) Nephropathy complicates approximately 30% of cases of type 1 and type 2 diabetes mellitus and is characterized clinically by proteinuria and progressive renal insufficiency. The typical glomerular lesion is glomerulosclerosis due to thickening of theGBM and expansion of the mesangium with extracellular matrix. Factors implicated as triggers for increased matrix production include glomerular hypertension; the direct effects of hyperglycemia on mesangial cells; advanced glycosylation end-products; growth factors such as growth hormone, insulin-like growth factor 1, and angiotensin II; cytokines such asTGF-b; hyperlipidemia; and cell sorbitol accumulation.

Complementary clinical and laboratory approaches suggest a central role for hemodynamic factors. Glomerular hydrostatic pressure andGFRincrease within months of the development of hyperglycemia. The mechanism by which diabetes mellitus induces glomerular hypertension is still being defined but appears to involve atrial natriuretic peptide. In this framework, glycosuria triggers increased reabsorption of glucose coupled to sodium in the proximal tubule, thereby increasing total-body sodium and extracellular fluid volume. As a compensatory response, atrial natriuretic peptide is released from cardiac myocytes and induces natriuresis in part by triggering afferent

arteriolar dilatation and thereby increasing intraglomerular pressure and GFR. Whereas this compensatory response is appropriate in the short term, sustained glomerular hypertension provokes thickening of theGBM, increased mesangial matrix production, and glomerulosclerosis and disruption of barrier function. In keeping with a central role for intra-glomerular pressure in the pathogenesis of diabetic nephropathy, angiotensin-converting enzyme inhibitors, which lower intraglomerular pressure, slow the progression of diabetic nephropathy, even in normotensive patients. It remains to be determined why diabetes mellitus and glomerular hypertension include glomerulosclerosis in some but not all individuals. Epidemiologic studies and studies of disease concordance in identical twins suggest that important, but as yet unidentified, genetic factors may play a role. It is likely that hemodynamic and metabolic factors act in concert to generate the final glomerulosclerotic phenotype in genetically predisposed patients.

Other Metabolic Diseases Several rare inherited lysosomal enzyme defects induce focal segmental glomerulosclerosis, probably by allowing accumulation of toxic metabolites in renal cells. *Fabry's disease* (a-galactosidase deficiency;Chap. 349) and *sialidosis* (*N*-acetylneuraminic acid hydrolase deficiency;Chap. 349) are the major culprits in this regard. Both tend to induce focal segmental or global glomerulosclerosis by preferentially affecting visceral epithelial cells, probably because these are terminally differentiated cells with a very slow replication rate. *Partial lipodystrophy* is a rare metabolic disorder characterized by lipoatrophy affecting the arms, neck, and chest, often with redistribution of fat to the hips and legs. Approximately one-third of patients develop glomerular disease, usually type II membranoproliferative glomerulopathy (dense deposit disease;Chap. 274).

HEMODYNAMIC GLOMERULAR INJURY

High intraglomerular pressure is a major cause of glomerular injury in humans and can result from systemic hypertension or a local change in glomerular hemodynamics (glomerular hypertension).

Systemic Hypertension (See also Chap. 246) Although the kidneys have evolved sophisticated mechanisms for autoregulating glomerular blood flow and pressure, marked or sustained increments in systemic blood pressure can overwhelm these compensatory systems and perturb glomerular morphology and function. In its most dramatic form, namely malignant hypertension, hemodynamic stress causes massive fibrinoid necrosis of afferent arterioles and glomeruli, thrombotic microangiopathy, acute renal failure, and a nephritic urinary sediment. Chronic sustained hypertension typically leads to arteriolar vasoconstriction and sclerosis, which, in turn, cause secondary atrophy and sclerosis of glomeruli and the tubulointerstitium. A variety of molecular signals appear to couple elevations in intravascular pressure to myointimal proliferation and eventually sclerosis of the vessel wall. These include growth factors such as angiotensin II, epidermal growth factor, andPDGF; cytokines such asTGF-b; and activation of stretch activated ion channels and early response genes.

Glomerular Hypertension The pathophysiology of diabetic nephropathy, discussed above, illustrates the importance of intraglomerular pressure as a stimulus for mesangial matrix production and glomerulosclerosis. Glomerular hypertension is also a key factor

in the pathogenesis of the progressive glomerulosclerosis and renal failure that complicate the adaptive response of remnant nephrons to increased workload following loss of the other nephrons from any cause, including chronic allograft failure (see below). Importantly, these changes in glomerular hemodynamics and pressure appear to precede the development of systemic hypertension and are independent risk factors for glomerular injury.

TOXIC GLOMERULOPATHIES

The renal microvasculature is a relatively uncommon site for toxic injury, by comparison with the tubular interstitium; however, there are a few important exceptions. Verotoxin, derived from *Escherichia coli* during bouts of infective diarrhea, is directly toxic to renal endothelium and induces the hemolytic-uremic syndrome. In this setting, verotoxin interacts with a specific cell membrane receptor, perturbs the antithrombotic phenotype of endothelium, and triggers the development of thrombotic microangiopathy. Irradiation, mitomycin, cyclosporine, and anovulants can also induce thrombotic microangiopathy through poorly defined mechanisms. Nonsteroidal anti-inflammatory drugs, rifampin, ampicillin, and interferon-a can induce an unusual combination of acute renal failure with nephrotic syndrome. The characteristic pathologic correlates of this syndrome are allergic interstitial nephritis and fusion of the foot processes of the visceral epithelial cells, the latter accounting for the marked proteinuria. How these structurally diverse agents induce epithelial cell injury is unclear.

DEPOSITION DISEASES

The glomerular deposition diseases are a group of diverse conditions in which abnormal proteins are deposited in glomeruli, where they provoke an inflammatory reaction and/or glomerulosclerosis. The major glomerular deposition diseases are cryoglobulinemia, amyloidosis, light and heavy chain deposition disease, and fibrillary/immunotactoid glomerulopathy. *Cryoglobulins* (Chap. 317) are immunoglobulins that precipitate in the cold and can be composed of either monoclonal immunoglobulin, usually generated by a lymphoproliferative malignancy (type I); a mixture of polyclonal immunoglobulin (usually IgG) and monoclonal immunoglobulin (usually IgM) directed to epitopes on polyclonal IgG (type II); or a mixture of polyclonal antibodies, one or more having anti-IgG activity (type III). As discussed above, cryoglobulins can induce nephritic-type and nephrotic-type injury depending on the rapidity, severity, and site of immunoglobulin deposition. Most cryoglobulinemic glomerulopathy is associated with type II cryoglobulins, the majority of which are now recognized to be triggered by chronic hepatitis B or C infection. *Glomerular amyloidosis* (Chaps. 275 and319) is one of the five most common causes of nephrotic syndrome in adults and is characterized by extracellular deposition of amyloid fibrils composed, in part, of fragments of immunoglobulin light chains (AL amyloid) or serum amyloid A, the acute-phase reactant (AA amyloid). In light chain deposition diseases, intact immunoglobulin light chains, usually kappa, are deposited in a granular, rather than fibrillary, pattern. The composition of the deposits in *fibrillary/immunotactoid glomerulopathy* is still being defined and may also include immunoglobulin and/or fibronectin-containing cryoglobulins. These different types of deposits, in addition to directly disrupting glomerular architecture, provoke mesangial matrix production and glomerulosclerosis. Fibrillary/immunotactoid glomerulopathy can also present as acute or subacute

glomerular inflammation. How these diverse deposits trigger glomerular matrix production and recruitment of inflammatory cells has yet to be determined.

INFECTIOUS CAUSES OF GLOMERULAR DISEASE

Infectious organisms can induce glomerular disease through several different mechanisms: (1) by direct infection of renal cells, (2) by elaborating nephrotoxins such as *E. coli*-derived verotoxin, (3) by inciting intraglomerular deposition of immune complexes (e.g., postinfectious glomerulonephritis) or cryoglobulins (e.g., hepatitis B or C), and (4) by providing a chronic stimulus for amyloid fibril formation, as in AA amyloidosis. Direct infection of glomerular cells is a relatively rare mechanism of injury but has been implicated in the pathogenesis of nephropathy associated with HIV. This entity is characterized histologically by an aggressive form of focal segmental glomerulosclerosis, microcystic tubular dilatation, and interstitial fibrosis. Viral genome and several proteins have been detected in glomerular and tubular cells in this disease, and infection of glomerular cells induces expression ofTGF-b, a major stimulus for mesangial matrix production and sclerosis.

INHERITED GLOMERULAR DISEASES

Alport's syndrome (hereditary nephritis; Chap. 275), the prototypical inherited glomerular disease, is usually transmitted as an X-linked dominant trait,although autosomal recessive forms have been reported. Patients afflicted with the classic X-linked form have a mutation in the COL4A5 gene that encodes thea5 chain of type IV collagen located on the X chromosome. As a result, theGBM is irregular with longitudinal layering, splitting, or thickening, and patients develop hematuria, progressive glomerulosclerosis, and renal failure. *Thin basement membrane disease* is another relatively common disorder of the GBM. In contrast to Alport's syndrome, this entity is usually inherited as an autosomal dominant or recessive trait and appears to be relatively benign. As the name suggests, the basement membrane is thin but otherwise ultrastructurally normal. Patients typically experience recurrent benign hematuria. The molecular basis for thin basement membrane disease has yet to be elucidated fully; however, defects in the gene encoding the a4 chain of type IV collagen have been reported in some families. Rarer hereditary glomerular diseases include *nail-patella syndrome* (osteoonychodysplasia), which is associated with a relatively benign mottling of the basement membrane with lucent rarefractions; *partial lipodystrophy*, which is associated with type II membranoproliferative glomerulonephritis (dense deposit disease); and *familial lecithin-cholesterol acyltransferase deficiency*, which is associated with distortion of the basement membrane by irregular rounded lucent zones, increased mesangial matrix production, and progressive sclerosis and renal insufficiency.

GLOMERULAR ADAPTATION TO NEPHRON LOSS

Nephron loss, from any cause, is followed by compensatory hyperfiltration in the remaining functional glomeruli. This adaptive response is appropriate in the short-term and maintainsGFR. Over years, however, the hyperfiltering remnant nephrons develop focal and segmental glomerulosclerosis, and eventually global sclerosis, that manifests clinically as proteinuria, hypertension, and progressive renal insufficiency. Sustained glomerular capillary hypertension has been implicated as a major stimulus for

glomerulosclerosis in this setting. Increased glomerular blood flow and ultrafiltration pressure are early findings in remnant nephrons in most experimental models in which the function of more than 50% of nephron mass has been lost through surgical ablation, immunologic or toxic injury, or other mechanisms. Sustained glomerular hypertension is thought to stimulate the accumulation of extracellular matrix by perturbing the function of visceral epithelial and mesangial cells, either directly or by increasing the flux of circulating macromolecules through the glomerular capillary wall. As with most forms of glomerulosclerosis,TGF-bmay be an important regulator of matrix accumulation in remnant nephrons. Angiotensin II,PDGF, and endothelins are other potential modulators of this process. Maneuvers that lower intraglomerular pressure, such as low-protein diet or treatment with angiotensin-converting enzyme inhibitors, slow the development of glomerulosclerosis and renal failure. Glomerular hypertrophy, intracapillary microthrombi, recruited macrophages, and hyperlipidemia are other potential stimuli for glomerulosclerosis. Indeed, glomerular capillary hypertension and hypertrophy appear to be independent risk factors that could act synergistically to cause progressive renal insufficiency. Intriguingly, angiotensin II may trigger TGF-b production in remnant nephrons, suggesting that angiotensin-converting enzyme inhibitors may be renoprotective through complementary effects on glomerular hemodynamics and matrix production.

(Bibliography omitted in Palm version)

Back to Table of Contents

274. THE MAJOR GLOMERULOPATHIES - *Hugh R. Brady*, *Yvonne M. O'Meara*, *Barry M. Brenner*

Glomerular injury can arise from diverse renal-limited and systemic diseases and is the major cause of end-stage renal disease (ESRD) requiring dialysis and transplantation. In this**chapter**, we describe the epidemiology, clinical presentations, pathology, and treatment of the major glomerulopathies. We focus on the *primary glomerulopathies*, glomerular diseases in which the pathologic process is confined to the kidney and in which systemic features are a direct consequence of impaired glomerular filtration (e.g., hypervolemia, hypertension, uremic syndrome). Considered here are the five major clinical presentations of glomerulopathy: acute nephritic syndrome, rapidly progressive glomerulonephritis (RPGN), nephrotic syndrome, asymptomatic abnormalities of the urinary sediment (hematuria, proteinuria), and chronic glomerulonephritis.**Glomerulopathies associated with systemic diseases (secondary glomerulopathies*) are discussed in Chap. 275. The nomenclature pertaining to the classification and clinicopathologic description of glomerular disease and the pathogenetic mechanisms of glomerular injury are reviewed in Chap. 273.

ACUTE NEPHRITIC SYNDROME AND RAPIDLY PROGRESSIVE GLOMERULONEPHRITIS

CLINICAL FEATURES AND CLINICOPATHOLOGIC CORRELATES

The *acute nephritic syndrome* is the clinical correlate of acute glomerular inflammation. In its most dramatic form, the acute nephritic syndrome is characterized by sudden onset (i.e., over days to weeks) of *acute renal failure* and *oliguria* (<400 mL of urine per day). Renal blood flow and glomerular filtration rate (GFR) fall as a result of obstruction of the glomerular capillary lumen by infiltrating inflammatory cells and proliferating resident glomerular cells. Renal blood flow and GFR are further compromised by intrarenal vasoconstriction and mesangial cell contraction that result from local imbalances of vasoconstrictor (e.g., leukotrienes, platelet-activating factor, thromboxanes, endothelins) and vasodilator substances (e.g., nitric oxide, prostacyclin) within the renal microcirculation. *Extracellular fluid volume expansion*, *edema*, and *hypertension* develop because of impaired GFR and enhanced tubular reabsorption of salt and water. As a result of injury to the glomerular capillary wall, urinalysis typically reveals *red blood cell casts*, dysmorphic red blood cells, leukocytes, and subnephrotic proteinuria of<3.5 g per 24 h ("nephritic urinary sediment"). *Hematuria* is often macroscopic.

The classic pathologic correlate of the nephritic syndrome is *proliferative glomerulonephritis*. The proliferation of glomerular cells is due initially to infiltration of the glomerular tuft by neutrophils and monocytes and subsequently to proliferation of resident glomerular endothelial and mesangial cells (endocapillary proliferation). In its most severe form, the nephritic syndrome is associated with acute inflammation of most glomeruli, i.e., *acute diffuse proliferative glomerulonephritis*. When less vigorous, fewer than 50% of glomeruli may be involved, i.e., *focal proliferative glomerulonephritis*. In milder forms of nephritic injury, cellular proliferation may be confined to the mesangium, i.e., *mesangioproliferative glomerulonephritis*.

RPGNis the clinical correlate of more *subacute glomerular inflammation*. Patients develop renal failure over weeks to months in association with a nephritic urinary sediment, subnephrotic proteinuria and variable oliguria, hypervolemia, edema, and hypertension. The classic pathologic correlate of RPGN is crescent formation involving most glomeruli (*crescentic glomerulonephritis*), crescents being half-moon-shaped lesions in Bowman's space composed of proliferating parietal epithelial cells and infiltrating monocytes (*extracapillary proliferation*). In practice, the clinical term *rapidly progressive glomerulonephritis* and the pathologic term *crescentic glomerulonephritis* are often used interchangeably. In addition to classic crescentic glomerulonephritis, in which crescents dominate the glomerular pathology, crescents can also develop concomitantly with proliferative glomerulonephritis or as a complication of membranous glomerulopathy and other more indolent forms of glomerular inflammation.

The acute nephritic syndrome andRPGN are part of a spectrum of presentations of immunologically mediated proliferative glomerulonephritis. Studies of experimental models suggest that nephritic syndrome and diffuse proliferative glomerulonephritis represent an acute immune response to a sudden large antigen load, whereas RPGN and crescentic glomerulonephritis represent a more subacute immune response to a smaller antigen load in presensitized individuals. At the other end of the spectrum, chronic low-grade immune injury presents with slowly progressive renal insufficiency or asymptomatic hematuria in association with focal proliferative or mesangioproliferative glomerulonephritis. These more indolent forms of immune-mediated glomerulonephritis are discussed later in this**chapter**.

ETIOLOGY AND DIFFERENTIAL DIAGNOSIS

Acute nephritic syndrome andRPGN can result from renal-limited *primary* glomerulopathy or from *secondary* glomerulopathy complicating systemic disease.Figure 274-1 highlights the histopathologic and serologic features that help distinguish among the major causes of nephritic syndrome and RPGN (see also Fig. 274-2). In general, rapid diagnosis and prompt treatment are critical to avoid the development of irreversible renal failure. Renal biopsy remains the "gold standard" for diagnosis. *Immunofluorescence microscopy* is particularly helpful and identifies three major patterns of deposition of immunoglobulin that define three broad diagnostic categories: (1) *granular* deposits of immunoglobulin, a hallmark of *immune-complex glomerulonephritis*; (2) *linear* deposition of immunoglobulin along the glomerular basement membrane (GBM), characteristic of anti-GBM disease; and (3) paucity or absence of immunoglobulin, so-called *pauci-immune glomerulonephritis* (Figs. 273-1 and274-2). Most patients (>70%) with full-blown acute nephritic syndrome have immune-complex glomerulonephritis. Pauci-immune glomerulonephritis is less common in this setting (<30%) and anti-GBM disease is rare (<1%). Among patients with RPGN, immune-complex glomerulonephritis and pauci-immune glomerulonephritis are equally prevalent (~45% each), whereas anti-GBM disease again accounts for a minority of cases (<10%).

Three *serologic markers* often predict the immunofluorescence microscopy findings in nephritic syndrome andRPGN and may obviate the need for renal biopsy in classic cases. They are the serum C3 level and titers of anti-GBMantibody and antineutrophil cytoplasmic antibody (ANCA) (Fig. 274-1). As discussed inChap. 273, the kidney is host

to immune attack in immune-complex glomerulonephritis, most cases being initiated either by in situ formation of immune complexes or less commonly by glomerular trapping of circulating immune complexes. These patients typically have hypocomplementemia (low C3 and CH_{50} in 90%) and negative anti-GBM and ANCA serology. The glomerulus is the direct target of immune attack in anti-GBM disease, glomerular inflammation being initiated by an autoantibody directed at a 28-kDa autoantigen on thea3 chain of type IV collagen. Approximately 90 to 95% of patients with anti-GBM disease have circulating anti-GBM autoantibodies detectable by immunoassay; serum complement levels are typically normal, and ANCA are usually not detected. The pathogenesis of pauci-immune glomerulonephritis is still being defined; however, most patients have circulating ANCA, implicating dysregulation of humoral immunity. The presence of mononuclear leukocytes in glomeruli and the paucity of glomerular immune deposits suggest that cellular mechanisms are also involved. Serum complement levels are typically normal, and anti-GBM titers are usually negative in ANCA-associated renal disease.

IMMUNE-COMPLEX GLOMERULONEPHRITIS

Immune-complex glomerulonephritis may (1) be idiopathic, (2) represent a response to a known antigenic stimulus (e.g., postinfectious glomerulonephritis), or (3) form part of a multisystem immune-complex disorder (e.g., lupus nephritis, Henoch-Schonlein purpura, cryoglobulinemia, bacterial endocarditis;Fig. 274-1). Here, we focus on *postinfectious glomerulonephritis*, the best characterized *primary* immune-complex glomerulonephritis. The major *secondary* immune-complex glomerulonephritides are discussed inChap. 275. Nephritic syndrome andRPGNoccasionally complicate two other primary glomerulopathies, namely, membranoproliferative glomerulonephritis (MPGN) and IgA nephropathy when there is a florid proliferative component. Because nephrotic syndrome and asymptomatic hematuria are more common presentations of MPGN and IgA nephropathy, respectively, these glomerulopathies are discussed in later sections on nephrotic syndrome and asymptomatic urinary abnormalities.

POSTSTREPTOCOCCAL GLOMERULONEPHRITIS

This is the prototypical postinfectious glomerulonephritis and a leading cause of acute nephritic syndrome. Most cases are sporadic, though the disease can occur as an epidemic. Glomerulonephritis develops, on average, 10 days after pharyngitis or 2 weeks after a skin infection (impetigo) with a nephritogenic strain of group Ab-hemolytic streptococcus. The known nephritic strains include M types 1, 2, 4, 12, 18, 25, 49, 55, 57, and 60. Immunity to these strains is type-specific and long-lasting, and repeated infection and nephritis are rare. Epidemic poststreptococcal glomerulonephritis is most commonly encountered in children of 2 to 6 years of age with pharyngitis during the winter months. This entity appears to be decreasing in frequency, possibly due to more widespread and prompt use of antibiotics. Poststreptococcal glomerulonephritis in association with cutaneous infections usually occurs in a setting of poor personal hygiene or streptococcal superinfection of another skin disease.

The classic clinical presentation of poststreptococcal glomerulonephritis is full-blown nephritic syndrome with oliguric acute renal failure; however, most patients have milder disease. Indeed, subclinical cases outnumber overt cases by four- to tenfold during

epidemics. Patients with overt disease present with gross hematuria (red or "smoky" urine), headache, and generalized symptoms such as anorexia, nausea, vomiting, and malaise. Swelling of the renal capsule can cause flank or back pain. Physical examination reveals hypervolemia, edema, and hypertension. The urinary sediment is nephritic, with dysmorphic red blood cells, red cell casts, leukocytes, occasionally leukocyte casts, and subnephrotic proteinuria. Fewer than 5% of patients develop nephrotic-range proteinuria. The latter may only manifest as acute nephritis resolves and renal blood flow andGFRrecover. Coexistent rheumatic fever is extremely rare.

The serum creatinine is often mildly elevated at presentation. Serum C3 levels and CH_{50} are depressed within 2 weeks in ~90% of cases. C4 levels are characteristically normal, indicating activation of the alternate pathway of complement. Complement levels usually return to normal within 6 to 8 weeks. Persistently depressed levels after this period should suggest another cause, such as the presence of a C3 nephritic factor (see "Membranoproliferative Glomerulonephritis"). The majority of patients (>75%) have transient hypergammaglobulinemia and mixed cryoglobulinemia. The antecedent streptococcal infection may still be evident or may have resolved either spontaneously or in response to antibiotic therapy. Most patients (>90%) have circulating antibodies against streptococcal exoenzymes such as antistreptolysin O (ASO), anti-deoxyribonuclease B (anti-DNAse B), antistreptokinase (ASKase), anti-nicotinyl adenine dinucleotidase (anti-NADase), and antihyaluronidase (AHase). ASO, anti-DNAse B, anti-NAD, and AHase are most useful after pharyngeal infection, whereas anti-DNAse B and AHase are more sensitive indices of streptococcal skin infection. Antibody titers tend to rise after 7 days, peak after 1 month, and return to normal levels after 3 to 4 months. These tests are relatively specific, with a false-positive rate of<5%. Early antibiotic therapy may prevent the development of an antibody response.

Acute poststreptococcal glomerulonephritis is usually diagnosed on clinical and serologic grounds, without resort to renal biopsy, especially in children with a typical antecedent history. The characteristic lesion on light microscopy is diffuse proliferative glomerulonephritis. Crescents are uncommon, and extraglomerular involvement is usually mild. Immunofluorescence microscopy reveals diffuse granular deposition of IgG and C3, giving rise to a "starry sky" appearance (Fig. 274-2). More extensive immunoglobulin deposition throughout the glomerular capillary wall ("garland pattern") is associated with a worse prognosis. The characteristic finding on electron microscopy is the presence of large electron-dense immune deposits in the subendothelial, subepithelial, and mesangial areas. The acute inflammatory reaction is initiated, in large part, by the subendothelial and mesangial deposits, which activate complement and trigger leukocyte recruitment and glomerular injury. Subepithelial deposits are often more prominent on electron microscopy, however, probably because the subendothelial and mesangial deposits are scavenged more efficiently by invading phagocytes. Extensive subepithelial immune deposits, or "humps," tend to be associated with worse proteinuria, being juxtaposed to the glomerular filtration barrier for protein.**The pathogenesis of immune-complex glomerulonephritis is discussed extensively in Chap. 273*.

In addition to poststreptococcal glomerulonephritis, the nephritic syndrome andRPGN can complicate acute immune-complex glomerulonephritis due to other viral, bacterial,

fungal, and parasitic infections. Several warrant specific mention. Diffuse proliferative immune-complex glomerulonephritis is a well-described complication of acute and subacute bacterial endocarditis and is usually associated with hypocomplementemia. The glomerular lesion typically resolves following eradication of the cardiac infection. *Shunt nephritis* is a syndrome characterized by immune-complex glomerulonephritis secondary to infection of ventriculoatrial shunts inserted for treatment of childhood hydrocephalus. The most common offending organism is coagulase-negative staphylococcus. Renal impairment is usually mild and associated with hypocomplementemia. Nephrotic syndrome complicates 30% of cases. Acute proliferative glomerulonephritis can also complicate *chronic suppurative infections* and *visceral abscesses*. Patients typically present with a fever of unknown origin and an active urine sediment. Although immune deposits containing IgG and C3 are detected on renal biopsy, serum complement levels are usually normal.

TREATMENT

Treatment of poststreptococcal glomerulonephritis focuses on eliminating the streptococcal infection with antibiotics and providing supportive therapy until spontaneous resolution of glomerular inflammation occurs. Patients are usually confined to bed during the acute inflammatory phase. Diuretics and antihypertensive agents are employed to control extracellular fluid volume and blood pressure. Dialysis is rarely needed to control hypervolemia or the uremic syndrome. Poststreptococcal glomerulonephritis carries an excellent prognosis and rarely causesESRD. Microscopic hematuria may persist for as long as 1 year after the acute episode but eventually resolves. Whereas complete recovery is the rule in children, adults may occasionally be left with residual renal impairment.

Antiglomerular Basement Membrane Disease Anti-GBMdisease is an autoimmune disease in which autoantibodies directed against type IV collagen induceRPGN and crescentic glomerulonephritis (Figs. 273-1,274-1, and274-2). Acute nephritic syndrome is rare. Between 50 and 70% of patients have lung hemorrhage; the clinical complex of anti-GBM nephritis and lung hemorrhage is referred to as *Goodpasture's syndrome*. Anti-GBM disease is a rare disorder of unknown etiology with an annual incidence of 0.5 per million. There is a bimodal peak in incidence. Patients with Goodpasture's syndrome are typically young males (5 to 40 years; male-female ratio of 6:1). In contrast, patients presenting during the second peak in the sixth decade rarely suffer lung hemorrhage and have an almost equal sex distribution. The target antigen is a component of the noncollagenous (NCI) domain of thea3 chain of type IV collagen, thea3 chain being preferentially expressed in glomerular and pulmonary alveolar basement membrane. The trigger(s) for loss of self-tolerance to this Goodpasture antigen has not been well defined. A genetic predisposition is suggested by an association with HLA-DRw2 and occasional occurrence in identical twins. Patients with lung hemorrhage are more likely to be cigarette smokers and to have suffered a recent upper respiratory tract infection or exposure to volatile hydrocarbon solvents. These observations suggest that diverse insults to the alveolar basement membrane may render previously sequestered Goodpasture antigens available for interaction with circulating autoantibodies. It is not clear whether environmental factors also trigger the onset of nephritis. Binding of anti-GBM antibodies to the GBM induces activation of complement, leukocyte recruitment, necrotizing proliferative glomerulonephritis, disruption of the glomerular

capillary wall, leakage of fibrin into Bowman's space, and crescent formation (Chap. 273). A similar sequence of events in the lung leads to disruption of the alveolar capillary wall and pulmonary hemorrhage.

Anti-GBMdisease commonly presents with hematuria, nephritic urinary sediment, subnephrotic proteinuria, and rapidly progressive renal failure over weeks, with or without pulmonary hemorrhage. When pulmonary hemorrhage occurs, it usually predates nephritis by weeks or months. Hemoptysis can vary from fluffy pulmonary infiltrates on chest x-ray and mild dyspnea on exertion to life-threatening pulmonary hemorrhage. Hypertension is unusual and occurs in fewer than 20% of cases.

The diagnostic serologic marker is circulating anti-GBMantibodies with a specificity for the NCI domain of thea3 chain of type IV collagen (Fig. 274-1). Anti-GBM antibodies are detected in the serum of >90% of patients with anti-GBM nephritis by specific immunoassay. If immunoassays are not available, circulating anti-GBM antibodies can be detected in 60 to 80% of patients by indirect immunofluorescence, i.e., by incubating the patient's serum with stored sections of normal human kidneys. Complement levels are normal. About 20% of patients have low titers ofANCA, usually a perinuclear ANCA (Chap. 275), the pathophysiologic significance of which is unclear. Occasional patients have a positive cytoplasmic ANCA, which may signal the presence of coexistent extraglomerular renal vasculitis. Patients with lung involvement frequently have microcytic, hypochromic, iron-deficiency anemia from alveolar hemorrhage, and abnormal bilateral hilar and basilar interstitial shadowing on chest x-ray that may be difficult to distinguish from pulmonary edema or infection. The diffusion capacity for carbon dioxide is a useful tool for distinguishing among the latter diagnoses, being increased in patients with lung hemorrhage due to uptake of carbon monoxide by alveolar blood, and reduced in patients with infection or pulmonary edema.

Renal biopsy is the gold standard for diagnosis of anti-GBMnephritis. The typical morphologic pattern on light microscopy is diffuse proliferative glomerulonephritis, with focal necrotizing lesions and crescents in >50% of glomeruli (crescentic glomerulonephritis). Immunofluorescence microscopy reveals linear ribbon-like deposition of IgG along the GBM (Fig. 274-2). C3 is present in the same distribution in 70% of patients. Prominent IgG deposition along the tubule basement membrane and tubulointerstitial inflammation is found occasionally. Electron microscopy reveals nonspecific inflammatory changes without immune deposits. Typical features on lung biopsy include alveolar hemorrhage, disruption of alveolar septa, hemosiderin-laden macrophages, and linear staining of IgG along the alveolar capillary basement membrane.

It should be noted that Goodpasture's syndrome is not the only cause of the pulmonary-renal syndrome (i.e., renal failure and lung hemorrhage). Other important causes of this clinical complex include severe cardiac failure complicated by pulmonary edema (often blood-tinged) and prerenal azotemia; renal failure from any cause complicated by hypervolemia and pulmonary edema; immune complex-mediated vasculitides such as systemic lupus erythematosus (SLE), Henoch-Schonlein purpura, and cryoglobulinemia; pauci-immune vasculitides such as Wegener's granulomatosis and polyarteritis nodosa; infections such as Legionnaire's disease; and renal vein thrombosis with pulmonary embolism. In general, these disorders can be differentiated

by astute analysis of the clinical, serologic, and histopathologic findings.

TREATMENT

Prior to the introduction of immunosuppressive therapy, greater than 80% of patients with anti-GBMnephritis developedESRDwithin 1 year, and many patients died from pulmonary hemorrhage or complications of uremia. With early and aggressive use of plasmapheresis, glucocorticoids, cyclophosphamide, and azathioprine, renal and patient survival have improved dramatically. In general, emergency plasmapheresis is performed daily or on alternate days until anti-GBM antibodies are not detected in the circulation (usually 1 to 2 weeks). Prednisone (1 mg/kg per day) is started simultaneously, in combination with either cyclophosphamide (2 to 3 mg/kg per day) or azathioprine (1 to 2 mg/kg per day) to suppress new synthesis of anti-GBM antibodies. The speed of initiation of therapy is a critical determinant of outcome. One-year renal survival approaches 90% if treatment is started before serum creatinine exceeds 442 umol/L (5 mg/dL) and falls to about 10% if renal failure is more advanced. Patients who require dialysis at presentation rarely recover renal function. Serial anti-GBM titers are monitored to gauge response to therapy. Relapses are not unusual and are often heralded by rising antibody titers. In patients with ESRD, renal transplantation is a viable treatment option. Recurrence of anti-GBM nephritis in the allograft is extremely unusual provided that anti-GBM antibody titers have been consistently negative for 2 to 3 months prior to transplantation. However, in occasional patients with Alport's syndrome, when the allograft presents normal GBM components to the immune system of the recipient for the first time, anti-GBM nephritis can occur de novo in renal allografts.

Pauci-Immune Glomerulonephritis The major pauci-immune glomerulonephritides are *idiopathic renal-limited crescentic glomerulonephritis*, *microscopic polyarteritis nodosa*, and *Wegener's granulomatosis* (Fig. 274-1).RPGN is a more common clinical presentation than acute nephritic syndrome, and the usual pathology is necrotizing glomerulonephritis with crescents affecting>50% of glomeruli (crescentic glomerulonephritis). The marked overlap of clinical features and glomerular histopathology, and the presence of circulatingANCA in most patients, suggest that these entities are a spectrum of a single disease. Here, we focus on idiopathic renal-limited crescentic glomerulonephritis.**The ANCA-associated glomerulopathies with extrarenal features, namely Wegener's granulomatosis, Churg-Strauss syndrome, and microscopic polyarteritis nodosa, are discussed in Chap. 275.*

Idiopathic Renal-Limited Crescentic Glomerulonephritis This is more common in middle-aged and older patients and shows a slight male preponderance. Patients typically present withRPGN, nephritic syndrome being rare.ANCA, usually a perinuclear ANCA IgG with specificity for myeloperoxidase (Chap. 275), are detected in 70 to 90% of patients (Fig. 274-1). The erythrocyte sedimentation rate and C-reactive protein levels may be elevated; however, C3 levels are typically normal, and circulating immune complexes, cryoglobulins, and anti-GBMantibodies are not detected. Most patients have crescents on light microscopy, often associated with necrotizing glomerulonephritis. Immune deposits are scanty or absent. Immunofluorescence microscopy reveals abundant fibrin deposits within crescents (Fig. 274-2). Most cases are treated aggressively with glucocorticoids, with or without cyclophosphamide or azathioprine (Chap. 275).

NEPHROTIC SYNDROME

GENERAL FEATURES AND COMPLICATIONS

The *nephrotic syndrome* is a clinical complex characterized by a number of renal and extrarenal features, the most prominent of which are proteinuria of>3.5 g per 1.73 m_2 per 24 h (in practice, >3.0 to 3.5 g per 24 h), hypoalbuminemia, edema, hyperlipidemia, lipiduria, and hypercoagulability. It should be stressed that the key component is *proteinuria*, which results from altered permeability of the glomerular filtration barrier for protein, namely theGBM and the podocytes and their slit diaphragms. The other components of the nephrotic syndrome and the ensuing metabolic complications are all secondary to urine protein loss and can occur with lesser degrees of proteinuria or may be absent even in patients with massive proteinuria.

In general, the greater the proteinuria, the lower the serum albumin level. *Hypoalbuminemia* is compounded further by increased renal catabolism and inadequate, albeit usually increased, hepatic synthesis of albumin. The pathophysiology of *edema* formation in nephrotic syndrome is poorly understood. The *underfilling hypothesis* postulates that hypoalbuminemia results in decreased intravascular oncotic pressure, leading to leakage of extracellular fluid from blood to the interstitium. Intravascular volume falls, thereby stimulating activation of the renin-angiotensin-aldosterone axis and the sympathetic nervous system and release of vasopressin (antidiuretic hormone), and suppressing atrial natriuretic peptide release. These neural and hormonal responses promote renal salt and water retention, thereby restoring intravascular volume and triggering further leakage of fluid to the interstitium. This hypothesis does not, however, explain the occurrence of edema in many patients in whom plasma volume is expanded and the renin-angiotensin-aldosterone axis is suppressed. The latter finding suggests that *primary renal salt and water retention* may also contribute to edema formation in some cases.

Hyperlipidemia is believed to be a consequence of increased hepatic lipoprotein synthesis that is triggered by reduced oncotic pressure and may be compounded by increased urinary loss of proteins that regulate lipid homeostasis. Low-density lipoproteins and cholesterol are increased in the majority of patients, whereas very low density lipoproteins and triglycerides tend to rise in patients with severe disease. Although not proven conclusively, hyperlipidemia may accelerate atherosclerosis and progression of renal disease.

Hypercoagulability is probably multifactorial in origin and is caused, at least in part, by increased urinary loss of antithrombin III, altered levels and/or activity of proteins C and S, hyperfibrinogenemia due to increased hepatic synthesis, impaired fibrinolysis, and increased platelet aggregability. As a consequence of these perturbations, patients can develop spontaneous *peripheral arterial or venous thrombosis*, *renal vein thrombosis*, and *pulmonary embolism.* Clinical features that suggest acute renal vein thrombosis include sudden onset of flank or abdominal pain, gross hematuria, a left-sided varicocele (the left testicular vein drains into the renal vein), increased proteinuria, and an acute decline inGFR. Chronic renal vein thrombosis is usually asymptomatic. Renal vein thrombosis is particularly common (up to 40%) in patients with nephrotic syndrome

due to membranous glomerulopathy, membranoproliferative glomerulonephritis, and amyloidosis.

Other metabolic complications of nephrotic syndrome include *protein malnutrition* and iron-resistant *microcytic hypochromic anemia* due to transferrin loss. *Hypocalcemia* and secondary hyperparathyroidism can occur as a consequence of vitamin D deficiency due to enhanced urinary excretion of cholecalciferol-binding protein, whereas loss of thyroxine-binding globulin can result in *depressed thyroxine levels*. An increased susceptibility to *infection* may reflect low levels of IgG that result from urinary loss and increased catabolism. In addition, patients are prone to unpredictable changes in the *pharmacokinetics* of therapeutic agents that are normally bound to plasma proteins.

ETIOLOGY AND DIFFERENTIAL DIAGNOSIS

Proteinuria>150 mg per 24 h is abnormal and can result from a number of mechanisms. *Glomerular proteinuria* results from leakage of plasma proteins through a perturbed glomerular filtration barrier; *tubular proteinuria* results from failure of tubular reabsorption of low-molecular-weight plasma proteins that are normally filtered and then reabsorbed and metabolized by tubular epithelium; *overflow proteinuria* results from filtration of proteins, usually immunoglobulin light chains, that are present in excess in the circulation. Tubular proteinuria virtually never exceeds 2 g per 24 h and thus, by definition, never causes nephrotic syndrome. Overflow proteinuria should be suspected in patients with clinical or laboratory evidence of multiple myeloma or other lymphoproliferative malignancy. Suspicion is heightened when there is a discrepancy between proteinuria detected by dipsticks, which are sensitive to albumin but not light chains, and the sulfosalicylic acid precipitation method, which detects both.

Nephrotic syndrome can complicate any disease that perturbs the negative electrostatic charge or architecture of theGBM and the podocytes and their slit diaphragms. Six entities account for greater than 90% of cases of nephrotic syndrome in adults: minimal change disease (MCD), focal and segmental glomerulosclerosis (FSGS), membranous glomerulopathy,MPGN, diabetic nephropathy, and amyloidosis. Diabetic nephropathy and amyloidosis, being manifestations of systemic diseases, are discussed inChap. 275. *Renal biopsy* is a valuable tool in adults with nephrotic syndrome for establishing a definitive diagnosis, guiding therapy, and estimating prognosis. Renal biopsy is not required in the majority of children with nephrotic syndrome as most cases are due to MCD and respond to empiric treatment with glucocorticoids.

MINIMAL CHANGE DISEASE

This glomerulopathy accounts for about 80% of nephrotic syndrome in children of younger than 16 years and 20% in adults (Table 274-1). The peak incidence is between 6 and 8 years. Patients typically present with nephrotic syndrome and benign urinary sediment. Microscopic hematuria is present in 20 to 30%. Hypertension and renal insufficiency are very rare.

MCD(also called nil disease, lipoid nephrosis, or foot process disease) is so named because glomerular size and architecture are normal by light microscopy. Immunofluorescence studies are typically negative for immunoglobulin and C3. Mild

mesangial hypercellularity and sparse deposits of C3 and IgM may be detected. Occasionally, mesangial proliferation is associated with scanty IgA deposits, similar to those found in IgA nephropathy. However, the natural history of this variant and response to therapy resemble classic MCD. Electron microscopy reveals characteristic *diffuse effacement of the foot processes of visceral epithelial cells* (Fig. 274-3). This morphologic finding is referred to as foot process fusion in the older literature.

The etiology ofMCD is unknown and the vast majority of cases are idiopathic (Table 274-1). MCD occasionally develops after upper respiratory tract infection, immunizations, and atopic attacks. Patients with atopy and MCD have an increased incidence of HLA-B12, suggesting a genetic predisposition. MCD, often in association with interstitial nephritis, is a rare side effect of nonsteroidal anti-inflammatory drugs (NSAIDs), rifampin, and interferon-a. The occasional association with lymphoproliferative malignancies (such as Hodgkin's lymphoma), the tendency for idiopathic MCD to remit during intercurrent viral infection such as measles, and the good response of idiopathic forms to immunosuppressive agents (see below) suggest an immune etiology. In children, the urine contains albumin principally and minimal amounts of higher molecular weight proteins such as IgG anda_2-macroglobulin. This *selective proteinuria* in conjunction with foot process effacement suggests injury to podocytes and loss of the fixed *negative charge* in the glomerular filtration barrier for protein. Proteinuria is typically nonselective in adults, suggesting more extensive perturbation of membrane permeability.

TREATMENT

MCDis highly steroid-responsive and carries an excellent prognosis. Spontaneous remission occurs in 30 to 40% of childhood cases but is less common in adults. Approximately 90% of children and 50% of adults enter remission following 8 weeks of high-dose oral glucocorticoids. In a typical regimen using prednisone, children receive 60 mg/m_2 of body surface area daily for 4 weeks, followed by 40 mg/m_2 on alternate days for an additional 4 weeks; adults receive 1 to 1.5 mg/kg body weight per day for 4 weeks, followed by 1 mg/kg per day on alternate days for 4 weeks. Up to 90% of adults enter remission if therapy is extended for 20 to 24 weeks. Nephrotic syndrome relapses in over 50% of cases following withdrawal of glucocorticoids. Alkylating agents are reserved for the small number of patients who fail to achieve lasting remission. These include patients who relapse during or shortly after withdrawal of steroids (steroid-dependent) and those who relapse more than three times per year (frequently relapsing). In these settings, cyclophosphamide (2 to 3 mg/kg per day) or chlorambucil (0.1 to 0.2 mg/kg per day) is started after steroid-induced remission and continued for 8 to 12 weeks. Cytotoxic agents may also induce remission in occasional steroid-resistant cases. These benefits must be balanced against the risk of infertility, cystitis, alopecia, infection, and secondary malignancies, particularly in children and young adults. Azathioprine has not been proven to be a useful adjunct to steroid therapy. Cyclosporine induces remission in 60 to 80% of patients; it is an alternative to cytotoxic agents and an option in patients who are resistant to cytotoxic agents. Unfortunately, relapse is usual when cyclosporine is withdrawn, and long-term therapy carries the risk of nephrotoxicity and other side effects. Long-term renal and patient survival is excellent in MCD.

FOCAL AND SEGMENTAL GLOMERULOSCLEROSIS WITH HYALINOSIS

The pathognomonic morphologic lesion inFSGS is sclerosis with hyalinosis involving portions (segmental) of fewer than 50% (focal) of glomeruli on a tissue section. The incidence of idiopathic (primary) FSGS has increased over the past two decades so that it now accounts for about one-third of cases of nephrotic syndrome in adults and as many as one-half of cases of nephrotic syndrome in blacks. FSGS can complicate a number of systemic diseases and sustained glomerular capillary hypertension following nephron loss from any cause (Table 274-2 andChap. 273).

IdiopathicFSGStypically presents as nephrotic syndrome (~66%) or subnephrotic proteinuria (~33%) in association with hypertension, mild renal insufficiency, and an abnormal urine sediment that contains red blood cells and leukocytes. Proteinuria is nonselective in most cases.

Light microscopy of renal biopsy tissue revealsFSGSwith entrapment of amorphous hyaline material, a process that shows a predilection for juxtamedullary glomeruli. The sclerotic scars contain areas of glomerular capillary collapse and hyaline material composed of collagen types I, III, and IV. Adhesions occur between areas of capillary collapse and Bowman's capsule. Immunofluorescence studies are usually negative. Electron microscopy reveals evidence of damage to visceral epithelial cells, including swelling and detachment of podocytes from theGBM, effacement of foot processes, transition to foam cells, and overt cell degeneration and necrosis.

The etiology of primaryFSGS is unclear (Table 274-2). There is evidence that a circulating nonimmunoglobulin permeability factor triggers FSGS in at least a subgroup of patients. The latter individuals tend to be young and prone to develop early recurrence of FSGS following renal transplantation. Plasmapharesis has been employed with variable success to control the nephrotic syndrome in this group. The overlap of clinical and morphologic features betweenMCD and FSGS has prompted some authorities to speculate that they are a spectrum of morphologic manifestations of a single pathogenetic process. FSGS is a potential long-term consequence of nephron loss from any cause. It can complicate congenital renal diseases such as congenital oligomeganephronia, in which both kidneys have a reduced complement of nephrons, and congenital unilateral agenesis. In addition, FSGS may develop following acquired loss of nephrons from extensive surgical ablation of renal mass; reflux nephropathy; glomerulonephritis; interstitial nephritis; sickle cell disease; and the combined effects of ischemia, cyclosporine nephrotoxicity, and rejection on renal allograft function (Table 274-2). It appears that>50% of nephrons must be lost for development of secondary FSGS.

TREATMENT

In contrast toMCD, spontaneous remission of primaryFSGS is rare and renal prognosis is relatively poor. Proteinuria remits in only 20 to 40% of patients treated with glucocorticoids for 8 weeks. Uncontrolled studies suggest that up to 70% respond when steroid therapy is prolonged for 16 to 24 weeks. Cyclophosphamide and cyclosporine, when used at doses described above for MCD, induce partial or complete remission in 50 to 60% of steroid-responsive patients but are generally ineffective in steroid-resistant

cases. Poor prognostic factors at presentation include hypertension, abnormal renal function, black race, and persistent heavy proteinuria. Renal transplantation is complicated by recurrence of FSGS in the allograft in about 50% of cases and graft loss in about 10%. Factors associated with an increased risk of recurrence include a short time interval between the onset of the FSGS andESRD, young age at onset, and possibly the presence of mesangial hypercellularity on renal biopsy.

MEMBRANOUS GLOMERULOPATHY

This lesion is a leading cause of idiopathic nephrotic syndrome in adults (30 to 40%) and a rare cause in children (<5%). It has a peak incidence between the ages of 30 to 50 years and a male-female ratio of 2:1 (Table 274-3). Membranous glomerulopathy derives its name from the characteristic light-microscopic appearance on renal biopsy, namely diffuse thickening of theGBM, which is most apparent upon staining with periodic acid-Schiff (PAS). Most patients (>80%) present with nephrotic syndrome, proteinuria usually being nonselective. Microscopic hematuria is present in up to 50% of cases, but red blood cells casts, macroscopic hematuria, and leukocytes are extremely rare. Hypertension is documented in only 10 to 30% of patients at the outset but is common later in patients with progressive renal failure. Serologic tests such as antinuclear antibody,ANCA, anti-GBM antibody, cryoglobulin titers, and complement levels are normal in the idiopathic form.

Light microscopy of renal biopsy sections reveals diffuse thickening of theGBMwithout evidence of inflammation or cellular proliferation. Silver staining demonstrates characteristic *spikes* along the GBM, which represent projections of new basement membrane engulfing subepithelial immune deposits. Immunofluorescence reveals granular deposition of IgG, C3, and the terminal components of complement (C5b-9) along the glomerular capillary wall. Electron-microscopic appearances vary depending on the stage of disease. The earliest finding is the presence of subepithelial immune deposits (Fig. 274-3). As these deposits enlarge, spikes of new basement membrane extend out between the immune deposits and begin to engulf them. With time, the deposits are completely surrounded and incorporated into the basement membrane.

The pathogenesis of idiopathic human membranous glomerulopathy is incompletely understood. The presence of electron-dense immune deposits that contain IgG and C3 suggest an immune process. About one-third of adult membranous nephropathy occurs in association with systemic diseases such asSLE, infections such as hepatitis B, malignancy, and drug therapy with gold and penicillamine (Table 274-3).

Nephrotic syndrome remits spontaneously and completely in up to 40% of patients with membranous glomerulopathy. The natural history of another 30 to 40% is characterized by repeated relapses and remissions. The final 10 to 20% suffer a slow progressive decline inGFR that typically culminates inESRDafter 10 to 15 years. Presenting features that predict a poor prognosis include male gender, older age, hypertension, severe proteinuria and hyperlipidemia, and impaired renal function. Controlled trials of glucocorticoids have failed to show consistent improvement in proteinuria or renal protection. Cyclophosphamide, chlorambucil, and cyclosporine have each been shown to reduce proteinuria and/or slow the decline in GFR in patients with progressive disease in small or uncontrolled studies. These observations need to be confirmed in

controlled prospective studies. Transplantation is a successful treatment option for patients who reach ESRD.

MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS

This morphologic entity, also known as mesangiocapillary glomerulonephritis, is characterized by thickening of theGBM and proliferative changes on light microscopy (Table 274-4). Two major types are identified; both are characterized by a diffuse increase in mesangial cellularity and matrix, and by thickening and reduplication of the GBM such that the lobular pattern of the glomerular tuft is exaggerated. The hallmark of type IMPGN is the presence of subendothelial and mesangial deposits on electron microscopy that contain C3 and IgG or IgM; rarely, IgA deposits are demonstrated by immunofluorescence microscopy (Fig. 274-3). The hallmark of type II MPGN (dense deposit disease) is the presence of electron-dense deposits within the GBM and other renal basement membranes (shown by electron microscopy) that stain for C3, but little or no immunoglobulin.

Most patients with type IMPGNpresent with heavy proteinuria or nephrotic syndrome, active urinary sediment, and normal or mildly impairedGFR. C3 levels are usually depressed, and C1q and C4 levels are borderline or low. Type I MPGN is an immune-complex glomerulonephritis and can be associated with a variety of chronic infections (e.g., bacterial endocarditis, HIV, hepatitis B and C), systemic immune-complex diseases (e.g.,SLE, cryoglobulinemia), and malignancies (e.g., leukemias, lymphomas). Type I MPGN is a relatively benign disease, and 70 to 85% of patients survive without clinically significant impairment of GFR. There is no proven therapy for patients with progressive disease beyond eradicating the underlying infection, malignancy, or systemic disease, when possible. The incidence of type I MPGN appears to be falling, possibly because the overall incidence of hepatitis C infection has fallen dramatically in western society over the past decade.

Type IIMPGN can also present with proteinuria and nephrotic syndrome; however, some patients present with nephritic syndrome,RPGN, or recurrent macroscopic hematuria. Type II MPGN is an autoimmune disease in which patients have an IgG autoantibody, termed *C3 nephritic factor*, that binds to C3 convertase, the enzyme that metabolizes C3, and renders it resistant to inactivation (Chap. 273). Type II MPGN runs a variable course; theGFRremains stable in some patients and declines gradually toESRDover 5 to 10 years in others. There is no effective therapy for this disease.

FIBRILLARY-IMMUNOTACTOID GLOMERULOPATHY

This emerging clinicopathologic entity accounts for 1% of diagnoses in most large renal biopsy series. Virtually all patients present with proteinuria, and>50% have nephrotic syndrome. The majority of patients also have hematuria, hypertension, and renal insufficiency. The light-microscopic appearances vary from mesangial expansion and basement membrane thickening withPAS-positive material to proliferative and crescentic glomerulonephritis. On electron microscopy, this PAS-positive material is observed to be composed of randomly arranged (fibrillary glomerulopathy) or organized bundles (immunotactoid glomerulopathy) of microfibrils and microtubules, the compositon of which has yet to be defined. The etiology of fibrillary-immunotactoid

glomerulopathy remains to be determined. Patients with the immunotactoid variant have an increased incidence of lymphoproliferative malignancy. There is no proven therapy for fibrillary-immunotactoid glomerulopathy, and many patients progress to ESRDover 1 to 10 years. Transplantation appears to be a viable option in the latter setting.

MESANGIAL PROLIFERATIVE GLOMERULONEPHRITIS

In 5 to 10% of patients with idiopathic nephrotic syndrome, renal biopsy reveals a diffuse increase in glomerular cellularity, predominantly due to proliferation of mesangial and endothelial cells, and infiltration by monocytes. Findings on immunofluorescence microscopy vary and include deposits of IgA, IgG, IgM, and/or complement, or absence of immune reactants. It is likely that this morphologic entity is, in fact, a heterogeneous group of diseases that includes atypical forms ofMCD andFSGS and milder or resolving forms of the immune-complex and pauci-immune glomerulopathies described above under nephritic syndrome andRPGN. In keeping with the heterogeneity of this diagnosis, the prognosis is variable. In general, persistent nephrotic-range proteinuria signals a poor prognosis, with many patients progressing toESRDover 10 to 20 years despite immunosuppressive therapy.

TREATMENT

Nephrotic Syndrome and Complications The treatment of nephrotic syndrome involves (1) specific treatment of the underlying morphologic entity and, when possible, causative disease (see above); (2) general measures to control proteinuria if remission is not achieved through immunosuppressive therapy and other specific measures; and (3) general measures to control nephrotic complications.

General measures may be warranted to control proteinuria in nephrotic syndrome if patients do not respond to immunosuppressive therapy and other specific measures and suffer progressive renal failure or severe nephrotic complications. Nonspecific measures that may reduce proteinuria include *angiotensin-converting enzyme (ACE) inhibitors*, and*NSAIDs*. The first of these measures aim to reduce proteinuria and slow the rate of progression of renal failure by lowering intraglomerular pressure and preventing the development of hemodynamically mediated focal segmental glomerulosclerosis. There is conclusive evidence that ACE inhibitors are renoprotective in human diabetic nephropathy (Chap. 275) and that ACE inhibitors slow the development of secondaryFSGS in experimental animals. Their role in the treatment of nephrotic syndrome in other settings is unproven. NSAIDs also reduce proteinuria in some patients with nephrotic syndrome, probably by altering glomerular hemodynamics andGBMpermeability characteristics. This potential benefit must be balanced against the risk of inducing acute renal failure, hyperkalemia, salt and water retention, and other side effects.

Complications of nephrotic syndrome that may require treatment include edema, hyperlipidemia, thromboembolism, malnutrition, and vitamin D deficiency. Edema should be managed cautiously by moderate *salt restriction*, usually 1 to 2 g/day, and the judicious use of *loop diuretics*. It is unwise to remove >1.0 kg of edema per day as more aggressive diuresis may precipitate intravascular volume depletion and prerenal azotemia. Administration of salt-poor albumin is not recommended as most is excreted

within 24 to 48 h. Whereas many nephrologists advocate lowering low-density lipoproteins and cholesterol levels with *lipid-lowering drugs* to prevent accelerated atherosclerosis and slow the rate of decline ofGFR, the value of such interventions in this setting has not been conclusively shown. *Anticoagulation* is indicated for patients with deep venous thrombosis, arterial thrombosis, and pulmonary embolism. Patients may be relatively resistant to heparin as a consequence of antithrombin III deficiency. Renal vein and vena caval angiography are probably indicated only when embolization occurs on anticoagulation and insertion of a caval filter is contemplated. There is no consensus regarding the optimal *diet* for patients with nephrotic syndrome. High-protein diets to prevent protein malnutrition are now in disfavor, since protein supplements have little, if any, effect on serum albumin levels and may hasten the progression of renal disease by increasing urinary protein excretion. The potential value of dietary protein restriction for reducing proteinuria must be balanced against the risk of contributing to malnutrition. *Vitamin D* supplementation is advisable in patients with clinical or biochemical evidence of vitamin D deficiency.

ASYMPTOMATIC ABNORMALITIES OF THE URINARY SEDIMENT

HEMATURIA

Most asymptomatic glomerular hematuria is due to *IgA nephropathy* (Berger's disease) or *thin basement membrane (TBM) disease* (benign hematuria). A rarer but more ominous cause of isolated hematuria is *Alport's syndrome*. The latter is the most common form of hereditary nephritis, is usually transmitted as an X-linked dominant trait, and is associated with sensorineural deafness, ophthalmologic abnormalities, and progressive renal insufficiency (Chap. 275). TBM disease is sometimes familial but, in contrast to Alport's syndrome, is usually a benign disorder. Asymptomatic hematuria may also be the presenting feature of indolent forms of most other primary and secondary proliferative glomerulopathies (Fig. 274-1). Glomerular hematuria must be distinguished from a variety of renal parenchymal and extrarenal causes of hematuria. It is particularly important to exclude malignancy of the kidney or urinary tract, particularly in older male patients (Chap. 94). Other potential diagnoses include vascular, cystic, and tubulointerstitial diseases; papillary necrosis; hypercalciuria and hyperuricosuria; benign prostatic hypertrophy; and renal calculi. Important clues to the presence of glomerular hematuria are the presence of urinary red blood cell casts, dysmorphic urinary red blood cells, proteinuria of greater than 2.0 g per 24 h, and clinical or serologic evidence of nephritic syndrome,RPGN, or a compatible systemic disease.

IgA Nephropathy (Berger's Disease) IgA nephropathy is the most common glomerulopathy worldwide and accounts for 10 to 40% of glomerulonephritis in most series (Table 274-5). The disease is particularly common in southern Europe and Asia and appears to be more common in blacks than whites. Familial clustering has been reported but is rare. No consistent HLA association has emerged, although HLA-B35 appears to be more common in French patients. Most cases are idiopathic. The renal and serologic abnormalities in IgA nephropathy and Henoch-Schonlein purpura (Chap. 275) are indistinguishable, and most authorities consider these to be a spectrum of a single disease. Less commonly, IgA nephropathy is found in association with systemic diseases, including chronic liver disease, Crohn's disease, gastrointestinal adenocarcinoma, chronic obstructive bronchiolitis, idiopathic interstitial pneumonia,

dermatitis herpetiformis, mycosis fungoides, leprosy, ankylosing spondylitis, relapsing polychondritis, and Sjogren's syndrome. In many of these conditions, IgA is deposited in the glomerulus without inducing inflammation, and this may be a clinically insignificant consequence of perturbed IgA homeostasis.

Patients with IgA nephropathy typically present with gross hematuria, often 24 to 48 h after a pharyngeal or gastrointestinal infection, vaccination, or strenuous exercise. Other cases are diagnosed upon detection of microscopic hematuria during routine physical examinations. Hypertension (20 to 30%) and nephrotic syndrome (~10%) are unusual at presentation. Light microscopy of renal biopsy specimens typically shows mesangial expansion by increased matrix and cells. Diffuse proliferation, cellular crescents, interstitial inflammation, and areas of glomerulosclerosis may be evident in severe cases. The diagnostic finding, for which the disease is named, is mesangial deposition of IgA, detected by immunofluorescence microscopy. C3 is usually detected in the area of immune deposits, and IgG is observed in 50% of cases. Electron microscopy reveals electron-dense deposits in the mesangium and, in severe cases, these extend into the paramesangial subendothelial space. The pathogenesis of IgA nephropathy is incompletely understood.

TREATMENT

There is no proven therapy for IgA nephropathy. A recent, relatively large randomized controlled trial suggested a benefit of fish oils in patients with progressive disease and heavy proteinuria; however, this experience has not been universal. Some authorities advocate a trial of high-dose glucocorticoids with or without cytotoxic agents in patients with severe nephrotic syndrome and those with nephritic syndrome orRPGN and evidence of active inflammation on renal biopsy.

IgA nephropathy typically smolders for decades, with patients often suffering exacerbations of hematuria and renal impairment during intercurrent infections. As many as 20 to 50% of patients developESRDwithin 20 years. Clinical predictors of a poor prognosis include older age, male sex, hypertension, nephrotic-range proteinuria, and renal insufficiency at presentation. Histologic features that predict an aggressive course include diffuse severe disease, extracapillary proliferation (crescents), extension of immune attack into the paramesangial subendothelial space, glomerulosclerosis, interstitial fibrosis, and arteriolar hyalinosis.

Thin Basement Membrane Disease (Benign Hematuria) This disorder can be heredofamilial or sporadic and is as common as IgA nephropathy in some series of asymptomatic hematuria. When familial, it is usually inherited as an autosomal dominant trait and is due to a defect in the gene encoding the a4 chain of type IV collagen.TBMdisease typically manifests in childhood as persistent hematuria. Intermittent hematuria and exacerbation of hematuria during upper respiratory tract infections have also been reported. The kidney is normal on light and immunofluorescence microscopy. TheGBM is thin (usually <275 nm in children and<300 nm in adults) by comparison with normal subjects. TBM disease is usually a benign condition, and progressive renal impairment or proteinuria should prompt a search for an alternative diagnosis. A small proportion of patients do, however, appear to develop hypertension and focal glomerulosclerosis upon long-term follow-up. The molecular

basis for the sporadic form of TBM disease has not been determined.

PROTEINURIA

Between 0.5 and 10% of the population have isolated proteinuria, defined as proteinuria in the presence of an otherwise normal urinary sediment, a radiologically normal urinary tract, and the absence of known renal disease. The majority of these patients excrete <2 g of protein per day, and more than 80% have an excellent prognosis (*benign isolated proteinuria*). A minority (10 to 25%) are found to have persistent proteinuria (*persistent isolated proteinuria*), some of whom develop progressive renal insufficiency over 10 to 20 years.

Benign Isolated Proteinuria The major categories of benign isolated proteinuria are idiopathic transient proteinuria, functional proteinuria, intermittent proteinuria, and postural proteinuria. *Idiopathic transient proteinuria* is usually observed in young adults and refers to dipstick-positive proteinuria in an otherwise healthy individual that disappears spontaneously by the next clinic visit. *Functional proteinuria* refers to transient proteinuria during fever, exposure to cold, emotional stress, congestive cardiac failure, or obstructive sleep apnea. This phenomenon is presumed to be mediated through changes in glomerular ultrafiltration pressure and/or membrane permeability. Patients with *intermittent proteinuria* have proteinuria in approximately half of their urine samples in the absence of other renal or systemic abnormalities. *Postural proteinuria* is proteinuria (usually<2.0 g per 24 h) that is evident only in the upright position. This disorder affects 2 to 5% of adolescents and may be transient (~80%) or fixed (~20%). Fixed postural proteinuria resolves within 10 to 20 years in most cases. In each of these conditions, renal biopsy reveals either normal renal parenchyma or mild and nonspecific changes involving podocytes or the mesangium. All carry an excellent prognosis.

Persistent Isolated Proteinuria Isolated proteinuria detected on multiple ambulatory clinic visits in both the recumbent and upright position usually signals a structural renal lesion. Virtually all glomerulopathies that induce nephrotic syndrome (see above) can cause persistent isolated proteinuria. The most common lesion on renal biopsy is mild mesangial proliferative glomerulonephritis with or without focal and segmental glomerulosclerosis (30 to 70%), followed by focal or diffuse proliferative glomerulonephritis (~15%) and interstitial nephritis (~5%). Although this clinical entity carries a worse prognosis than benign isolated proteinuria, the prognosis is still relatively good, with only 20 to 40% of patients developing renal insufficiency after 20 years. Furthermore, progression to ESRDis extremely rare. It is wise to exclude monoclonal gammopathy by urinary electrophoresis in older patients.

CHRONIC GLOMERULONEPHRITIS

This syndrome is characterized by persistent proteinuria and/or hematuria and renal insufficiency that progresses slowly over years. Chronic glomerulonephritis usually comes to light (1) upon routine urinalysis, (2) when routine blood tests reveal unexplained anemia or elevated blood urea nitrogen and creatinine, (3) following discovery of bilateral small kidneys on abdominal imaging, (4) during evaluation for secondary causes of hypertension, or (5) during a clinical exacerbation of glomerulonephritis triggered by pharyngitis (synpharyngitic) or other infections. Chronic

glomerulonephritis can be a manifestation of virtually all of the major glomerulopathies. Renal biopsy typically reveals a variable combination of proliferative, membranous, and sclerotic changes, depending on the causative glomerulopathy. Arteriosclerosis, induced by secondary hypertension, is a common finding in the renal vasculature. Tubulointerstitial inflammation and scarring are frequent additional findings and portend a poor prognosis. Glomerular hypertension and hyperfiltration through remnant functioning nephrons can hasten progression toESRD(Chap. 273). Treatment is directed at lowering systemic and glomerular hypertension, usually with anACEinhibitor, and controlling extracellular fluid volume, anemia, metabolic abnormalities, and the uremic syndrome through judicious use of diuretics, erythropoietin, and dietary modification (Chap. 270). Some patients develop ESRD and require renal replacement therapy with dialysis or transplantation.

(Bibliography omitted in Palm version)

Back to Table of Contents

275. GLOMERULOPATHIES ASSOCIATED WITH MULTISYSTEM DISEASES - *Yvonne M. O'Meara, Hugh R. Brady, Barry M. Brenner*

An array of multisystem diseases can cause glomerular injury, with glomerulopathy being either the dominant presenting feature or a relatively benign and clinically insignificant manifestation that is overshadowed by involvement of other organs. Glomerulopathies associated with multisystem diseases are often classified as *secondary* glomerulopathies to distinguish them from the *primary* glomerulopathies (Chap. 274) in which the pathology is limited to the kidneys. It should be emphasized, however, that most morphologic patterns of glomerular injury (see Table 273-1) can manifest as a renal-limited process (i.e., primary) or as part of a systemic disease (i.e., secondary). The diagnostic approach to glomerular disease involves identifying the presenting clinical syndrome (e.g., nephritic, nephrotic), defining the pathologic features (e.g., proliferative, crescentic, membranous), and attempting to establish the specific disease that provoked glomerular injury [e.g., systemic lupus erythematosus (SLE), Henoch-Schonlein purpura].

In this**chapter**, we focus on the epidemiology, clinicopathologic features, and management of the major glomerulopathies associated with systemic diseases.**The pathogenesis of glomerular injury is discussed in Chap. 273, and the overall place of the major glomerulopathies in the differential diagnosis of the major renal syndromes is described in Chap. 274.*

DIABETIC NEPHROPATHY (See also Chaps. 273 and 333)

Diabetic nephropathy is the leading cause of end-stage renal disease (ESRD) in western societies and accounts for 30 to 35% of patients on renal replacement therapy in North America. Type 1 diabetes mellitus (type 1 DM; formerly, insulin-dependent diabetes mellitus) and type 2 diabetes mellitus (type 2 DM; formerly, non-insulin-dependent diabetes mellitus) affect 0.5 and 4% of the population, respectively. Nephropathy complicates 30% of cases of type 1 DM and approximately 20% of cases of type 2 DM. However, most diabetic patients with ESRD have type 2 DM because of the greater prevalence of type 2 DM worldwide (90% of all individuals with diabetes). Risk factors for the development of diabetic nephropathy include hyperglycemia, systemic hypertension, glomerular hypertension and hyperfiltration, proteinuria, and possibly cigarette smoking, hyperlipidemia, and gene polymorphisms affecting the activity of the renin-angiotensin-aldosterone axis. For reasons that are unclear, ESRD from diabetic nephropathy is more common in blacks with type 2 DM than in whites (4:1 ratio), whereas the reverse is true for type 1 DM.

The pathophysiology, clinical features, and morphology of diabetic nephropathy are similar in type 1 and type 2DM, although the time course may be condensed in type 2 DM. Glomerular hypertension and hyperfiltration are the earliest renal abnormalities in experimental and human diabetes and are observed within days to weeks of diagnosis. Microalbuminuria, so named because the abnormal albumin excretion of 30 to 300 mg/24 h is below the limits of detection of standard dipsticks, develops after approximately 5 years of sustained glomerular hypertension and hyperfiltration in type 1 DM. Microalbuminuria is the first manifestation of injury to the glomerular filtration barrier and predicts the development of overt nephropathy. Dipstick-positive proteinuria,

ultimately reaching nephrotic levels, typically develops 5 to 10 years after the onset of microalbuminuria (i.e., 10 to 15 years after the onset of diabetes) and is associated with hypertension and progressive loss of renal function. In addition, patients can display features of tubulointerstitial disease such as hyperkalemia and type IV renal tubular acidosis.ESRDtypically develops 5 to 10 years after the development of overt nephropathy. As noted above, the course of diabetic nephropathy may be shorter in type 2 DM, and many patients present with established nephropathy and hypertension. Diabetic nephropathy is usually diagnosed on clinical grounds without a renal biopsy. Supportive clues are the presence of normal sized or enlarged kidneys, evidence of proliferative diabetic retinopathy, and a bland urinary sediment. Retinopathy is found in 90 and 60% of patients with type 1 and type 2 DM, respectively, who develop nephropathy.

The earliest morphologic abnormalities in diabetic nephropathy are thickening of the glomerular basement membrane (GBM) and expansion of the mesangium due to accumulation of extracellular matrix. With time, matrix accumulation becomes diffuse and is evident as eosinophilic, periodic acid Schiff-positive glomerulosclerosis on renal biopsy. Prominent areas of nodular matrix expansion (nodular glomerulosclerosis, the classic Kimmelsteil-Wilson lesion) are often superimposed on this background. The glomeruli and kidneys are typically normal or increased in size, distinguishing diabetic nephropathy from most other forms of chronic renal insufficiency (renal amyloidosis and polycystic kidney disease being other important exceptions). Immunofluorescence microscopy may reveal deposition of IgG along the GBM in a linear pattern, but this does not appear to be immunopathogenetic as in anti-GBM disease. Immune deposits are not seen. The renal vasculature typically displays evidence of atherosclerosis, as a consequence of hyperlipidemia, and hypertensive arteriosclerosis.

TREATMENT

Therapy is aimed at retarding the progression of nephropathy through control of blood sugar, systemic blood pressure, and glomerular capillary pressure. Glycemic control is achieved through regulation of diet and administration of oral hypoglycemic agents and insulin (Chap. 333). Angiotensin-converting enzyme (ACE) inhibitors are the drugs of choice as they control both systemic hypertension and intraglomerular hypertension by inhibiting the actions of angiotensin II on the systemic vasculature and renal efferent arterioles. ACE inhibitors also attenuate the stimulatory effect of angiotensin II on glomerular cell growth and matrix production. Because ACE inhibitors have been shown conclusively to delay the time toESRD by 50% in patients with type 1DM in a large randomized controlled trial and to delay progression significantly in type 2 DM, it is felt that all patients with diabetes should receive an ACE inhibitor on the development of microalbuminuria, even in the absence of systemic hypertension (Fig. 275-1). However, approximately 80% of patients with diabetes require more than one drug to control systemic hypertension, and aggressive lowering of blood pressure in these patients retards not only the rate of progression of nephropathy, but also the rate of progression of other complications of DM.

Diabetic nephropathy is the most common cause ofESRDrequiring renal replacement therapy, and patients with diabetes have the highest annual mortality rate (20 to 30%) of any group on dialysis, in large part as a result of accelerated atherosclerosis. The

survival rates of younger patients undergoing either peritoneal dialysis or hemodialysis are comparable; however, older patients with diabetes appear to have a higher mortality rate on peritoneal dialysis. Transplantation is the preferred mode of renal replacement therapy in patients who are otherwise medically suitable.

IMMUNOLOGICALLY MEDIATED MULTISYSTEM DISEASES

The glomerulus is a frequent target of injury in a variety of immunologically mediated multisystem diseases, particularly systemic vasculitis andSLE. Systemic vasculitis is usually classified according to the size of the inflamed vessel (Chap. 317). The major *large vessel* vasculitides are Takayasu's disease and giant cell arteritis. Glomerular injury is exceedingly rare in these diseases.

CLASSIC POLYARTERITIS NODOSA(See alsoChap. 317)

The typical glomerular lesion in classic polyarteritis nodosa (PAN) is ischemic collapse and obsolescence. Characteristic clinical and serologic features are hypertension, a bland urine sediment with subnephrotic proteinuria, slowly progressive renal insufficiency, normal serum complement levels, and absence of antineutrophil cytoplasmic antibodies (ANCA). Treatment with glucocorticoids and immunosupressive agents, such as cyclophosphamide, affords a 5-year patient survival of approximately 80%, as compared with 10% in untreated cases.

ANCA-ASSOCIATED SMALL-VESSEL VASCULITIS

The ANCA-associated small-vessel vasculitides are Wegener's granulomatosis, the microscopic and Churg-Strauss variants of polyarteritis nodosa, and pauci-immune renal-limited glomerulonephritis. These diseases share a number of clinocopathologic and serologic features and may represent a spectrum of manifestations of a single disease. They are more common in whites and older patients (mean age 57 years) and show a slight male preponderance. Their incidence peaks in the winter months, and many patients have a viral-like prodrome, suggesting a pathogenetic role for an infective agent. Patients typically present with nonspecific constitutional symptoms and signs such as lethargy, malaise, anorexia, weight loss, fever, arthralgias, and myalgias. Nonspecific laboratory abnormalities include a rapid sedimentation rate, elevated C-reactive protein, leukocytosis, thrombocytosis, and normochromic, normocytic anemia. Serum complement levels are typically normal.

The majority of patients with these conditions have circulatingANCA. It is not clear whether ANCA are involved in the pathogenesis of vasculitis or merely represent an epiphenomenon of the vasculitic process, and there is debate whether serial ANCA titers are useful for monitoring disease activity and predicting relapse. Some ANCA activate cytokine-primed neutrophils in vitro and provoke them to injure endothelial cells, suggesting a pathogenetic role (Chap. 273). ANCA are not entirely specific for vasculitis and are also found, albeit at low titers, in some patients (~20%) with anti-GBMdisease and in patients with inflammatory bowel disease, primary biliary cirrhosis, and other autoimmune disorders.

Patients withANCA-associated renal disease usually present with a nephritic urine

sediment and moderate proteinuria. Renal dysfunction can vary from a mild decrement in glomerular filtration rate (GFR) to rapidly progressive glomerulonephritis (RPGN). Renal biopsy typically reveals focal, segmental, necrotizing glomerulonephritis with crescent formation. Immunofluorescence and electron microscopy are remarkable for the paucity or absence of immunoglobulin, complement, and immune deposits: so-called pauci-immune glomerulonephritis. These findings are in stark contrast to the prominent granular deposition of IgG and C3 in immune-complex glomerulonephritides such as Henoch-Schonlein purpura and lupus nephritis. ANCA-associated vasculitis is usually responsive to combined therapy with glucocorticoids and cyclophosphamide, and the 5-year patient survival rate usually exceeds 75%. The distinguishing features of individual ANCA-associated renal diseases are summarized below.

Wegener's Granulomatosis (See also Chap. 317) Renal injury occurs in 80% of patients with Wegener's granulomatosis and varies from indolent smoldering inflammation to rapidly progressive renal failure. Cytoplasmic ANCA are detected at presentation in 80% of patients with renal disease and in 10% more on follow-up. Renal biopsy typically reveals focal, segmental, necrotizing pauci-immune glomerulonephritis with crescent formation. In contrast to the lung, granulomas are rarely seen in the kidney.

TREATMENT

Glucocorticoids and cyclophosphamide are the mainstays of treatment and dramatically ameliorate glomerular injury. Steroids are usually administered initially by pulse intravenous therapy on three consecutive days, followed by a daily oral dose of about 1 mg/kg body weight tapered to zero over 3 to 6 months. Cyclophosphamide is typically administered orally at a daily dose of 1 to 2 mg/kg or as monthly intravenous pulses of 1 g/m_2 of body surface area. Plasmapheresis may be a useful adjunct in patients with severe nephritis requiring dialysis. As many as 30% of patients relapse after treatment-induced remission. A persistently elevated or rising ANCA titer may predict relapse in individual patients; however, this relationship is not strong enough to justify treatment based on titers alone. Recent studies demonstrate that administration of trimethoprim-sulfamethoxazole reduces the relapse rate, possibly by eradicating nasal carriage of *Staphylococcus aureus*. Dialysis and renal transplantation afford excellent survival in patients with ESRD. Recurrence of Wegener's granulomatosis in the allograft is rare. ACE inhibitors may help to slow the progression to end-stage renal failure.

Microscopic Polyarteritis Nodosa (See also Chap. 317) Microscopic PAN is a systemic disease characterized by leukocytoclastic vasculitis involving multiple organ systems including the lungs, skin, joints, and kidneys. Clinical renal disease ranges from a nephritic urinary sediment with mild impairment of GFR to RPGN. The usual histopathologic lesion is pauci-immune focal segmental necrotizing and crescentic glomerulonephritis. Circulating ANCA are detected in 70 to 80% of patients at presentation, with cytoplasmic and perinuclear ANCA being equally prevalent. The treatment of microscopic PAN involves glucocorticoids and cyclophosphamide, as for Wegener's granulomatosis. Plasmapheresis may benefit patients with severe acute renal failure or massive pulmonary hemorrhage.

Churg-Strauss Syndrome (See also Chap. 317) Clinical renal involvement in

Churg-Strauss syndrome is relatively infrequent and usually limited to mild proteinuria and hematuria. Evolution to chronic renal failure is rare. Renal biopsy most frequently reveals extraglomerular pathology, with involvement of the renal vasculature and tubulointerstitium by granulomatous vasculitis. Focal segmental glomerulonephritis with crescents is also seen. A minority of patients have focal segmental necrotizing glomerulonephritis.

Henoch-Schonlein Purpura (See also Chap. 317) Extrarenal features of Henoch-Schonlein purpura include a petechial rash on the extremities, arthropathy, and abdominal pain. Nephritis is present in 80% of patients and manifests as a nephritic urine sediment and moderate proteinuria. Macroscopic hematuria and nephrotic-range proteinuria are uncommon. Light-microscopic appearances can vary from mild mesangial proliferation and expansion to diffuse proliferation with glomerular crescents. The glomerular lesion is identical to that found in IgA nephropathy (Berger's disease; seeChap. 274), suggesting that Henoch-Schonlein nephritis and IgA nephropathy are a spectrum of manifestations of a single disease. The *sine qua non* for diagnosis is the presence of mesangial IgA deposition on immunofluorescence microscopy. IgG and C3 are also detected. Electron microscopy reveals mesangial immune deposits. Immune complexes may also be present in the peripheral glomerular capillary wall and paramesangial areas. Biopsy of involved skin reveals dermal IgA deposition and leukocytoclastic vasculitis. IgA deposition is also seen in areas of uninvolved skin.

TREATMENT

Since there is no proven therapy for Henoch-Schonlein nephritis, treatment is supportive. Steroids and/or cytotoxic agents are often tried in patients with severe disease, but without compelling scientific evidence to support their use. The disease typically undergoes clinical exacerbations and remissions in the first year and then enters long-term remission. The prognosis is generally excellent; chronic renal failure and persistent hypertension occur in fewer than 10% of patients.

ESSENTIAL MIXED CRYOGLOBULINEMIA (See alsoChap. 317)

Renal involvement is most common with the mixed cryoglobulinemias (types II and III), which are more common in females and usually begin in the sixth decade. Most patients present with a variable combination of leukocytoclastic vasculitis, skin ulcerations, arthralgias, fatigue, and Raynaud's phenomenon. Renal disease is a complication in 50% of patients and usually develops after 12 to 24 months. The typical clinical renal manifestations are nephrotic-range proteinuria, microscopic hematuria, and hypertension. Acute nephritic syndrome occurs in 20 to 30%, and oliguric acute renal failure in about 5% of patients with renal disease. The characteristic morphologic lesions are diffuse mesangial proliferative or membranoproliferative glomerulonephritis. The glomerular capillaries frequently contain eosinophilic hyaline "pseudothrombi" composed of precipitated immunoglobulins. Granular deposition of IgG, IgM, and C3 is usually prominent on immunofluorescence microscopy. Electron microscopy typically reveals subendothelial deposits containing microfibrils and microtubules that display a characteristic "thumbprint" appearance.

Circulating levels of C3, C4, and CH50 are depressed in about 80% of patients with

renal involvement, and a transient antinuclear antibody (ANA) (speckled pattern) is sometimes detected. Abnormal liver function tests are found in about 15% of patients at presentation and in up to 50% subsequently. It now appears that most patients with essential mixed cryoglobulinemia (EMC) (i.e., idiopathic) are chronically infected with hepatitis C virus (HCV). In keeping with a pathogenetic role for this virus, HCV RNA has been isolated from the serum of patients with EMC, indicating active infection, and anti-HCV antibodies have been detected in both the serum and cryoprecipitates in association with viral antigens.

TREATMENT

Traditionally, glucocorticoids, with or without cyclophosphamide, and plasmapheresis were the standard treatment forEMC. Recent reports indicate that interferon a controls viral replication and stabilizes renal function in most patients infected with EMC andHCV. Unfortunately, most patients relapse when interferon ais discontinued, a major problem given the prohibitive cost of the drug. In general, patient and renal survival are good in EMC, with 75% of patients being alive at 10 years.

SYSTEMIC LUPUS ERYTHEMATOSUS (See also Chap. 311)

Renal involvement is clinically evident in 40 to 85% of patients withSLE; it varies from isolated abnormalities of the urinary sediment to full-blown nephritic or nephrotic syndrome or chronic renal failure. Most glomerular injury is triggered by the formation of immune complexes within the glomerular capillary wall; however, thrombotic microangiopathy may be the dominant reason for renal dysfunction in a small subset of patients with the antiphospholipid antibody syndrome.

Immune-Complex-Mediated Lupus Nephritis The renal biopsy has proven very useful for identifying the different patterns of immune-complex glomerulonephritis inSLE, which are diverse, portend different prognoses, and do not necessarily correlate with the clinical findings. Indeed, clinically silent lupus nephritis is well described in which the urinalysis is virtually normal but renal biopsy demonstrates varying degrees of injury.

The World Health Organization categorizes lupus nephritis into six histologic classes. *Class I* consists of a normal biopsy on light microscopy with occasional mesangial deposits on immunofluorescence microscopy. Patients in this category usually do not have clinical renal disease. Patients with *class II* or mesangial lupus nephritis have prominent mesangial deposits of IgG, IgM, and C3 on immunofluorescence and electron microscopy. Mesangial lupus nephritis is designated as class IIA when the glomeruli are normal by light microscopy and class IIB when there is mesangial hypercellularity. Microscopic hematuria is common with this lesion, and 25 to 50% of patients have moderate proteinuria. Nephrotic syndrome is not seen, and renal survival is excellent (>90% at 5 years). *Class III* describes focal segmental proliferative lupus nephritis with necrosis or sclerosis affecting fewer than 50% of glomeruli. Up to one-third of patients have nephrotic syndrome, and glomerular filtration is impaired in 15 to 25%. In *class IV* or diffuse proliferative lupus nephritis, most glomeruli show cell proliferation, often with crescent formation. Other features on light microscopy include fibrinoid necrosis and "wire loops," which are caused by basement membrane thickening and mesangial interposition between basement membrane and endothelial cells. Deposits of IgG, IgM,

IgA, and C3 are evident by immunofluorescence, and crescents stain positive for fibrin. Electron microscopy reveals numerous immune deposits in mesangial, subepithelial, and subendothelial locations. Tubuloreticular structures are frequently seen in endothelial cells. These are not specific for lupus nephritis and also occur in HIV-associated nephropathy. Electron microscopy may also reveal curvilinear parallel arrays of microfibrils, measuring approximately 10 to 15 nm in diameter, with "thumbprinting," similar to those seen in cryoglobulinemia. Nephrotic syndrome and renal insufficiency are present in at least 50% of patients with class IV disease. Diffuse proliferative lupus nephritis is the most aggressive renal lesion inSLE, and as many as 30% of these patients progress to terminal renal failure. *Class V* is termed membranous lupus nephritis because of its similarity to idiopathic membranous glomerulopathy. Thickening of theGBM is evident by light microscopy. Electron microscopy reveals predominant subepithelial deposits in addition to subendothelial and mesangial deposits. Proliferative changes may also be evident, but the predominant pattern is that of membranous glomerulopathy. Most patients present with nephrotic syndrome (90%), but significant impairment ofGFR is relatively unusual (10%). Tubulointerstitial changes such as active infiltration by inflammatory cells, tubular atrophy, and interstitial fibrosis are seen to varying degrees in lupus nephritis and are most severe in classes III and IV, especially in patients with long-standing disease. *Class VI* probably represents the end stages of proliferative lupus nephritis and is characterized by diffuse glomerulosclerosis and advanced tubulointerstitial disease. These patients are often hypertensive, may have nephrotic syndrome, and usually have impaired GFR.

Transformation from one class to another is relatively frequent. For example, class III often progresses to class IV spontaneously, and class IV can transform to class II or class V after treatment. Class II and class V lupus nephritis may predate other manifestations of lupus, whereas class III and class IV usually occur in patients who have systemic features ofSLE. A semiquantitative analysis can be performed by using a variety of features on renal biopsy, scored 0 to 3+, to derive indices of disease activity and chronicity. Features that suggest active inflammation include endocapillary proliferation, glomerular leukocyte infiltration, wire loop deposits, cellular crescents, and interstitial inflammation. In contrast, features that suggest chronicity include glomerulosclerosis, fibrous crescents, tubular atrophy, and interstitital fibrosis. In some, but not all, studies, these indices have been useful in predicting response to therapy and renal prognosis.

Patients with active lupus nephritis have a range of serologic abnormalities. Hypocomplementemia is present in 75 to 90% of patients and is most striking with diffuse proliferative glomerulonephritis.ANA are usually detected (95 to 99%), although not specific forSLE. ANA titers tend to fall with treatment, and ANA may not be detected during remissions. Anti-double-stranded DNA (dsDNA) antibodies are highly specific for SLE, and changes in their titers correlate with the activity of lupus nephritis. Almost 100% of patients taking procainamide and 65% of patients taking hydralazine develop ANA; however, overt lupus, including nephritis, occurs in fewer than 10% of these patients, and anti-DNA antibodies are not usually detected. Other antibodies found in patients with SLE include anti-Sm (17 to 30%; highly specific, but not sensitive); anti-RNP, which frequently accompanies anti-Sm in low titer; anti-Ro (35%); anti-La (15%); and anti-histone antibodies (70% of patients with SLE and 95% of patients with drug-induced lupus).

TREATMENT

The treatment of lupus nephritis is controversial and based largely on the class of injury and disease activity. Because there is relatively poor correlation between clinical features (urinalysis findings, serum creatinine) and histologic class, the renal biopsy findings are an important guide to therapy. Treatment is not indicated for class I and most cases of class II lupus nephritis, as these histologic patterns portend an excellent prognosis (100% and>90% 5-year survival rates, respectively). Extrarenal manifestations may warrant treatment with glucocorticoids, salicylates, or antimalarials. Glucocorticoids and cyclophosphamide are the mainstays of therapy for patients with proliferative nephritis (classes III and IV). High-dose steroids given as intravenous boluses (pulse therapy) are usually effective at rapidly controlling acute glomerular inflammation. Cyclophosphamide and azathioprine are important adjuncts to steroid therapy and appear to afford better long-term preservation of renal function than steroids alone. Intravenous pulse cyclophosphamide is as efficacious as oral therapy and appears to be less toxic. Most authorities advocate an initial regimen of monthly intravenous boluses of cyclophosphamide for 6 months. Subsequent therapy is tailored to disease activity and typically involves dosing every 3 to 6 months for a total treatment period of 18 to 24 months. The initial dose of cyclophosphamide is 0.5 g/m_2, and the dose is increased gradually to a maximum of 1 g/$m_2$unless patients develop leukopenia or other side effects. Steroids are usually started simultaneously at 1 mg/kg per day and are tapered over the first 6 months to a maintenance dose of 5 to 10 mg/d for the duration of cyclophosphamide therapy. Five-year renal survival rates of 60 to 90% have been obtained with this and similar regimens. A large randomized, prospective trial indicated that plasmapheresis does not offer additional benefit in patients with severe proliferative lupus nephritis. Mycophenolate mofetil has recently been used to treat patients with lupus nephritis that is resistant to steroids and cyclophosphamide.

The management of membranous lupus nephritis is less well defined. As with idiopathic membranous glomerulopathy, the incidence of spontaneous remission approaches 50% in membranous lupus nephritis, and the course of the disease is generally indolent, with a 70- to 90% renal survival rate at 5 years. Some authorities advocate steroids andACEinhibitors at the time of diagnosis, whereas others reserve them for patients with progressive renal insufficiency or severe nephrotic syndrome. Useful parameters for monitoring the response to therapy and predicting relapse include the activity of the urine sediment, proteinuria,GFR, serum complement levels, and anti-dsDNA titers. Despite maximal immunosuppressive therapy, about 20% of patients with aggressive lupus nephritis developESRDrequiring dialysis.SLEtends to become quiescent with advanced uremia, and patients rarely develop systemic flares once they commence dialysis. Recurrence of nephritis and systemic flares are also very uncommon after renal transplantation, and allograft survival rates are comparable to those in patients with other causes of ESRD.

ANTIPHOSPHOLIPID ANTIBODY SYNDROME AND THROMBOTIC MICROANGIOPATHY

Patients with this syndrome can develop a variable degree of renal impairment due to thrombotic microangiopathy. The latter typically affects the interlobular arteries,

arterioles, and glomerular capillaries and is characterized by intravascular microthrombi and swelling of endothelial cells. Decreased levels of tissue plasminogen activator and increased level of a_2-antiplasmin, both of which would tend to promote thrombosis, have been described in this syndrome. Anticoagulation to maintain the Internalized Normal Ratio (INR)>3.0 may be beneficial in reducing the incidence of recurrent thromboses. There are uncontrolled reports of a benefit of plasmapheresis in the setting of acute renal failure secondary to thrombotic microangiopathy.

RHEUMATOID ARTHRITIS (See also Chap. 312)

Although extra-articular manifestations are present in 35% of patients with rheumatoid arthritis, direct involvement of the kidney by rheumatoid disease is rare, and glomerular injury is usually secondary to amyloid A (AA) amyloidosis or a side effect of drug therapy. AA amyloidosis is a complication experienced by 10 to 20% of patients with rheumatoid arthritis, and renal involvement is evident clinically in 3 to 10% of these patients (nephrotic syndrome, renal insufficiency). Amyloidosis is more frequent in patients with rheumatoid arthritis of long duration (>10 years), with circulating rheumatoid factor, and with destructive arthropathy. Less frequent glomerular lesions include mesangial proliferative glomerulonephritis and basement membrane thickening by subepithelial immune deposits. Gold and penicillamine may cause nephrotic syndrome by inducing membranous glomerulopathy, whereas nonsteroidal anti-inflammatory drugs (NSAIDs) can trigger the nephrotic syndrome by inducing minimal change nephropathy, usually in association with acute interstitial nephritis (see below).

SJOGREN'S SYNDROME (See alsoChap. 314)

Tubulointerstitial injury is the most common form of renal involvement in Sjogren's syndrome and usually presents as either Fanconi's syndrome, distal renal tubular acidosis, or impairment of renal concentrating ability. Glomerulonephritis is relatively rare and should prompt a search for evidence of secondary causes. Membranous glomerulopathy and membranoproliferative glomerulonephritis (MPGN) are the most common lesions. Anecdotal reports describe successful therapy with glucocorticoids and cytotoxic agents.

POLYMYOSITIS AND DERMATOMYOSITIS (See also Chap. 382)

Occasional cases of focal mesangial proliferative glomerulonephritis with mesangial deposition of IgG and complement have been described in polymyositis/dermatomyositis. Membranous glomerulopathy has also been reported, particularly when polymyositis/dermatomyositis is associated with malignancy.

MIXED CONNECTIVE TISSUE DISEASE (See also Chap. 313)

Mixed connective tissue disease is a syndrome that includes features ofSLE, scleroderma, and polymyositis and is associated with high titers of antiribonucleoprotein antibodies and negative antismooth muscle antibodies. Renal involvement occurs in fewer than 15% of patients and manifests as hematuria and subnephrotic proteinuria. The usual pathologic lesion is membranous glomerulopathy orMPGN. The prognosis is

usually excellent, and steroid therapy may be useful in rare patients with progressive renal disease.

GLOMERULAR DEPOSITION DISEASES

The glomerular deposition diseases are characterized by deposition of abnormal proteins, usually immunoglobulins or fragments thereof, within the glomerulus. They include amyloidosis, light and heavy chain deposition disease, cryoglobulinemia, and fibrillary/immunotactoid glomerulonephritis. Here, we focus on amyloidosis and light chain deposition disease (LCDD). Cryoglobulinemic nephropathy is described above in the discussion on systemic vasculitis.**Fibrillary and immunotactoid glomerulopathy are discussed in Chap. 274.*

AMYLOIDOSIS (See alsoChap. 319)

Amyloidosis is classified according to the major component of its fibrils: for example, immunoglobulin light chains in amyloid L (AL) amyloidosis, serum amyloid A in AA amyloidosis,b_2-microglobulin in dialysis-associated amyloidosis, and amyloid bprotein in Alzheimer's disease and Down's syndrome. Amyloid deposits also contain a nonfibrillar component called the P component, a seruma$_1$glycoprotein with a high affinity for the fibrillar components of all forms of amyloid. AL and AA amyloidosis frequently involve the kidneys, whereas involvement by other forms of amyloidosis is very rare.

There is substantial overlap in the renal clinicopathologic presentations ofAL andAAamyloidosis. Glomeruli are involved in 75 to 90% of patients, usually in association with involvement of other organs. The clinical correlate of glomerular amyloid deposition is nephrotic-range proteinuria. In addition, over 50% of patients have impaired glomerular filtration at diagnosis. Hypertension is present in about 20 to 25%. Renal size is usually normal or slightly enlarged. A minority of patients present with renal failure due to amyloid deposition in the renal vasculature or with Fanconi's syndrome, nephrogenic diabetes insipidus, or renal tubular acidosis due to involvement of the tubulointerstitium. Rectal biopsy and abdominal fat pad biopsy reveal amyloid deposits in about 70% of patients and may obviate the need for renal biopsy.

Renal biopsy gives a very high yield if there is clinical evidence of renal involvement. The earliest pathologic changes are mesangial expansion by amorphous hyaline material and thickening of theGBM. Further amyloid deposition results in the development of large nodular eosinophilic masses. When stained with Congo red, these deposits show apple-green birefringence under polarized light. Immunofluorescence microscopy is usually only weakly positive for immunoglobulin light chains because amyloid fibrils are usually derived from the variable region of light chains. Electron microscopy reveals the characteristic nonbranching extracellular amyloid fibrils of 7.5 to 10 nm in diameter. Tubulointerstitial and vascular deposits of amyloid are also seen and may occasionally be more prominent than glomerular deposits.

TREATMENT

Most patients with renal involvement byALamyloidosis developESRDwithin 2 to 5 years. No treatment has been shown consistently to improve this prognosis; however, some

success has been reported with a combination of melphalan and prednisone. Preliminary studies have reported a benefit of high-dose melphelan with autologous stem cell transplantation. Colchicine delays the onset of nephropathy in patients with familial Mediterranean fever but has not proved useful in patients with established disease or with other forms of amyloid. Remissions may be achieved inAAamyloidosis by eradication of the underlying cause. Renal replacement therapy is offered to patients who reach ESRD; however, the 1-year survival rate on dialysis is low (~66%) by comparison with other causes of ESRD. Most patients die from extrarenal complications, particularly cardiovascular disease. Renal transplantation is a viable option in patients with AA amyloidosis whose primary disease has been eradicated. Transplantation is also an option for patients with AL amyloidosis, although a poor prognosis because of extrarenal organ involvement may preclude them as candidates. Here again, the survival rate is lower by comparison with other causes of ESRD; most of the excess mortality is due to infectious and cardiovascular complications. Recurrence of amyloidosis in the allograft is common but rarely leads to graft loss.

LIGHT CHAIN DEPOSITION DISEASE (See also Chap. 113)

Renal involvement is a complication in 90% of patients withLCDD and is often the dominant feature. Nephrotic syndrome and renal impairment are the usual presenting features. Microscopic hematuria occurs in about 20% of patients. Defective hydrogen ion and potassium excretion and urinary concentration may be evident if light chains are deposited predominantly in the tubules. The most common pathologic lesion on renal biopsy is ribbon-like thickening of the tubular basement membrane due to light chain deposition. Mesangial expansion and nodular glomerulosclerosis are found in about 33% of patients. This light-microscopic appearance resembles that of idiopathicMPGN and diabetic nephropathy. Superimposed crescentic change is occasionally seen. Immunofluorescence studies are strongly positive for monoclonal light chains, in contrast toALamyloid, because the constant region of the immunoglobulin is typically deposited. The tissue deposits in LCDD are granular rather than fibrillar on electron microscopy, appear more amorphous in character, do not stain with Congo red, and seem to have a greater affinity for basement membranes.

The prognosis ofLCDD is poor when it is associated with multiple myeloma, and most patients progess rapidly toESRD. Treatment with melphalan and prednisone has been reported to reduce proteinuria and stabilize renal function in uncontrolled studies. In the absence of myeloma, the prognosis is somewhat more variable, and several patients have undergone successful renal transplantation.

WALDENSTROM'S MACROGLOBULINEMIA (See alsoChap. 113)

This disorder is characterized by monoclonal proliferation of an IgM-secreting clone of plasma cells. The circulating IgM paraprotein frequently gives rise to the hyperviscosity syndrome, which may compromise renal blood flow andGFR. Direct renal involvement is rare and, when present, involves deposition of large amorphous deposits of eosinophilic material in the glomerular capillaries. Renal amyloidosis can also occur.

DRUG-INDUCED GLOMERULAR DISEASE

A variety of drugs damage the glomerular filtration barrier and induce proteinuria and nephrotic syndrome. In contrast, drug-induced proliferative glomerulonephritis is rare. The more common drug-induced glomerulopathies are discussed here. Additional associations are included in Table 275-1 andTable 71-1.

NSAIDshave a variety of renal side effects, including hemodynamically mediated acute renal failure, salt and water retention, hyponatremia, hyperkalemia, papillary necrosis, acute interstitial nephritis, nephrotic syndrome, andESRD. Nephrotic syndrome and acute renal failure frequently coexist due to a combination of acute interstitial nephritis and a glomerular lesion that is identical to that of minimal change disease. This entity occurs most commonly in patients on propionic acid derivatives such as fenoprofen, ibuprofen, and naproxen but can occur with other NSAIDs, ampicillin, rifampin, and interferona. Withdrawal of the drug usually results in resolution of renal disease. Membranous nephropathy is also described as an idiosyncratic reaction to NSAIDs.

Gold therapy, administered by injection or orally, induces proteinuria in 5 to 25% of patients with rheumatoid arthritis. Proteinuria develops after 4 to 6 months of therapy, and up to 33% of patients develop full-blown nephrotic syndrome. Renal biopsy typically reveals membranous glomerulopathy, though minimal change disease or mesangial proliferative lesions have also been described. Progressive renal impairment is rare. Nephrotic syndrome is more common in patients who are HLA-B8/DR3 positive, suggesting a genetic susceptibility. Withdrawal of the drug leads to gradual resolution of the proteinuria.

Penicillamine also induces proteinuria in 5 to 30% of patients. As with gold, the underlying glomerular lesion is usually membranous glomerulopathy, and proteinuria gradually resolves after withdrawal of the drug. Acute renal failure secondary to crescentic glomerulonephritis with immune deposits has also been described.

Intravenous heroin use is associated with an increased incidence of focal and segmental glomerulosclerosis (heroin-associated nephropathy). It is not clear whether the nephrotoxin in this setting is heroin itself or a contaminant. Heroin-associated nephropathy occurs predominantly in blacks and is characterized by nephrotic syndrome, hypertension, and a gradual progression toESRDover a period of 3 to 5 years. The pathologic features are similar to those of idiopathic focal segmental glomerulosclerosis, although mesangial deposition of IgM and C3 may be more prominent. The incidence of this disease appears to be declining steadily. Potential reasons for the decline include increased purity of street heroin and a bias to attribute focal segmental glomerulosclerosis to HIV infection when both risk factors coexist. Intravenous *amphetamine* abuse is a rare cause of systemic necrotizing vasculitis.

HEREDITARY DISEASES WITH GLOMERULAR INVOLVEMENT

ALPORT'S SYNDROME (See also Chap. 351)

Alport's syndrome is the most common hereditary nephritis and is usually transmitted as an X-linked dominant trait. The genetic defect resides in the gene for thea5 chain of type IV collagen located on the long arm of the X chromosome; type IV collagen is a major structural component of theGBM. Numerous genetic mutations have been

detected, ranging from major deletions to point mutations, and this genetic heterogeneity is reflected in the phenotypic variations of the disease. In the X-linked forms, males usually present with microscopic hematuria, proteinuria (nephrotic-range in 30%), and progressive renal insufficiency. Common extrarenal manifestations include sensorineural hearing loss (~60%), bilateral anterior lenticonus (~15 to 30%), and recurrent corneal erosions. Platelet defects are described but are rare. Female carriers usually have mild disease and do not develop renal insufficiency. Autosomal dominant and recessive forms also exist in which there are mutations in the gene for the a3 chain of type IV collagen, and males and females are equally affected. Genetic analysis to detect mutations in the genes encoding the a3 and a5 chains of type IV collagen may become the diagnostic method of choice.

Typical light-microscopic features on renal biopsy include mesangial hypercellularity, focal and segmental glomerulosclerosis, chronic tubulointerstitial fibrosis, atrophy, and accumulation of foam cells. Electron microscopy reveals thickening, fragmentation, and lamellation of the lamina densa of theGBM. Patchy thinning of the GBM may also be seen, especially early in the course of the disease and in female carriers.

Males with the disease tend to progress toESRD and are suitable candidates for dialysis and transplantation. About 5% of transplant recipients develop anti-GBMdisease in the renal allograft; their immune system recognizes normal GBM of the transplanted kidney as a foreign antigen. These patients can have antibodies against the a3 (Goodpasture antigen) ora5 chains of type IV collagen, probably because defective synthesis of the a5 chain results in defective incorporation or orientation of the a3 chain in the GBM.

SICKLE CELL DISEASE (See alsoChap. 106)

Glomerular disease is common (15 to 30%) in homozygotes for sickle cell disease. Glomerular hyperfiltration and hypertrophy occur within the first 5 years of life. Approximately 15 to 30% of patients develop proteinuria in the first three decades, and 5% developESRD. The glomerular pathology is usually focal segmental glomerulosclerosis, probably due to sustained glomerular capillary hypertension.MPGN is also seen on occasion. Predictors of chronic renal failure are worsening anemia, proteinuria, nephrotic syndrome, and hypertension.ACEinhibitors may slow the progression of renal disease by lowering systemic and glomerular capillary hypertension.

FABRY'S DISEASE (See also Chap. 349)

In patients with Fabry's disease, renal biopsy reveals accumulation of neutral glycosphingolipids with terminal a-galactosyl moieties in lysosomes of glomerular, tubular, vascular, and interstitial cells. Focal and global glomerulosclerosis are later features. Electron microscopy reveals stacked, concentric lamellar profiles known as "myeloid" bodies, which are characteristic. Renal disease manifests in the late teens to early twenties with lipiduria, proteinuria with minimal hematuria, nephrotic syndrome, hypertension, and progressive renal insufficiency. The most striking systemic manifestations are skin lesions (angiokeratomas), corneal and lens opacities, painful dysesthesias of the extremities, and arthropathy of the terminal interphalangeal joints. The diagnosis of Fabry's disease can often be made by careful physical examination,

especially if many of the typical clinical features are present. Measurement of urinary glycosphingolipids and estimation of peripheral leukocytea-galactosidase levels help confirm the diagnosis. The renal lesion is progressive, and these patients often tolerate hemodynamic changes during dialysis poorly because of progressive vascular disease. Successful renal transplantation has been reported despite recurrence in the allograft.

NAIL-PATELLA SYNDROME

The nail-patella syndrome is a rare hereditary disorder transmitted as an autosomal dominant trait. The abnormal gene is located on the long arm of chromosome 9, and candidate genes include the a1 chain of type V collagen and the LIM homeodomain protein Lmxlb. The phenotype is characterized by multiple osseous abnormalities, primarily affecting the elbows and knees, and nail dysplasia. About 50% of patients have clinically evident nephropathy. The light-microscopic features on renal biopsy include localGBMthickening, tubular atrophy, interstitial fibrosis, and varying degrees of glomerular sclerosis. Electron microscopy reveals irregular thickening of the GBM, with electrolucent areas giving it a "moth-eaten" appearance. Cross-striated fibrils with the periodicity of collagen can be identified in the mesangium and basement membrane. The disease usually manifests clinically as asymptomatic hematuria and proteinuria, occasionally in the nephrotic range, but it may be silent. The renal lesion is relatively benign, and progression toESRDoccurs in 10 to 30% of patients.

LIPODYSTROPHY

MPGNtype II (dense deposit disease) is the most frequent glomerular lesion in patients with lipodystrophy (80%), whereas MPGN type I affects the remainder (20%). The disease occurs mostly in females between the ages of 5 and 15 years, and the clinical presentation and course are similar to those of idiopathic MPGN, namely nephrotic-range proteinuria and progressive renal insufficiency. Low C3 levels are common in association with C3 nephritic factor (Chap. 274).

LECITHIN-CHOLESTEROL ACYLTRANSFERASE DEFICIENCY (See alsoChap. 344)

Renal manifestations of this disease include proteinuria, microscopic hematuria, and progressive renal insufficiency. Renal biopsy typically reveals focal and segmental glomerulosclerosis. Electron-microscopic findings include irregular rounded, lucent lacunae that contain solid or laminated dense structures in theGBM, mesangial matrix, and Bowman's capsular and renal tubular basement membranes. Endothelial cell detachment is also evident, and capillary lumens may be occluded by vacuolated foam cells. Recurrence of the disease has been documented in the renal allograft but without marked impairment of graft function.

GLOMERULAR LESIONS ASSOCIATED WITH INFECTIOUS DISEASES

VIRAL INFECTIONS

Hepatitis B, hepatitis C, and HIV are strongly associated with glomerular disease (Table 275-2). Glomerular lesions associated with *hepatitis B virus* (HBV) infection include membranous glomerulopathy,MPGN, IgA nephropathy, essential mixed

cryoglobulinemia, and polyarteritis nodosa. Membranous glomerulopathy is most common. In endemic areas, such as Asia and Africa, 80 to 100% of children and 30 to 45% of adults with membranous glomerulopathy have HBV surface antigenemia. HBV antigens have been identified in renal immune deposits, suggesting in situ immune-complex formation after planting of HBV antigens or trapping of circulating immune complexes containing HBV antigens. Patients typically present with nephrotic syndrome and microscopic hematuria. Hypertension and renal impairment are rare. The most common associated hepatic lesion is chronic persistent or chronic active hepatitis. In nonendemic areas there is a male preponderance, and many patients are intravenous drug users or have other risk factor for acquisition of HBV. The asymptomatic carrier state of HBV is frequently associated with MPGN in endemic areas. Hypertension and azotemia are more common with this morphologic pattern than with membranous glomerulopathy. Children with HBV-associated membranous glomerulonephritis have a good prognosis, and almost two-thirds enter spontaneous remission within 3 years.ESRD is rare. In contrast, 30% of adults develop progressive renal failure within 5 years, with 10% reaching ESRD. Steroids and cytotoxic agents are contraindicated as they may lead to increased viral replication and worsening of liver disease. Interferon a may reduce proteinuria and stabilize renal function in patients with progressive disease.

HCVinfection (Chap. 295) should be considered in all patients with cryoglobulinemic proliferative glomerulonephritis,MPGN, and membranous glomerulopathy. These three clinocopathologic entities may represent a spectrum of morphologic manifestations of the same pathogenetic process, namely HCV-induced immune-complex disease. Up to 30% of patients with chronic HCV infection have an abnormal urinary sediment. HCV infection accounts for 10 to 20% of type I MPGN and is a major cause of essential mixed cryoglobulinemia. Renal biopsy reveals typical features of type I MPGN and IgG, IgM, C3, and/or cryoglobulin deposits. Most patients present with nephrotic syndrome and microscopic hematuria and may have red blood cell casts. Liver function tests are usually abnormal, and C3 levels are typically depressed. Anti-HCV antibodies are detected in most patients, and viral RNA has been documented in blood and cryoglobulins. Various treatments have been reported to be useful in HCV-induced renal disease including steroids, cytotoxic agents, and plasmapheresis; however, controlled trials to support their use are lacking. Interferon a has been demonstrated to clear antigenemia, lower cryoglobulin levels, and stabilize renal disease. Unfortunately, relapse is usual once the drug is discontinued.

HIV infection (Chap. 309) has been associated with focal segmental glomerulosclerosis, acute diffuse proliferative glomerulonephritis, and mesangioproliferative glomerulonephritis, including IgA nephropathy,MPGN, and membranous glomerulopathy. The classic and most common HIV-associated glomerulopathy is an aggressive form of focal segmental glomerulosclerosis, an entity that is termed *HIV-associated nephropathy (HIVAN)*. This disease may be the first manifestation of infection in otherwise asymptomatic patients. HIVAN is more common in blacks than in other ethnic groups and is more frequent in intravenous drug abusers with HIV infection than in homosexuals. The disease has been described in all high-risk groups, however, including infants of HIV-positive mothers. Renal biopsy typically reveals visceral epithelial cell swelling, collapse of the glomerular capillary tuft, severe tubulointerstitial inflammation, and microcystic dilatation of renal tubules. Electron microscopy

characteristically reveals severe visceral epithelial cell injury and tubuloreticular inclusions in glomerular endothelial cells, tubular cells and infiltrating leukocytes. This constellation of findings has been termed collapsing glomerulopathy, but it should be emphasized that a similar picture can be seen in the absence of HIV infection. The presence of tubuloreticular inclusions and the aggressive clinical course distinguish HIVAN from idiopathic focal segmental glomerulosclerosis. The mechanisms of renal cell injury are still being defined. Viral DNA has been demonstrated in the renal epithelia of HIV-infected patients with and without nephropathy, suggesting that pathogenetic factors, other than infection of cells, are required for induction of disease. The typical clinical correlates of HIVAN are severe nephrotic syndrome and rapid progression toESRD, occurring in weeks to months. Despite early reports of poor survival of patients on dialysis, more recent studies indicate improved survival for both asymptomatic patients with HIV and patients with full-blown AIDS. There is no proven therapy for HIVAN. The initial experience with combined highly active antiretroviral therapy (triple therapy) suggests that these regimens have reduced the incidence of nephropathy in HIV-infected patients and improved prognosis in patients with established nephropathy.

BACTERIAL INFECTIONS (Table 275-2)

Immune-complex glomerulonephritis is a relatively frequent complication of inefective *endocarditis* (Chap. 126). Other mechanisms of renal injury in bacterial endocarditis include embolic renal infarction, septic abscesses, acute tubular necrosis secondary to septicemia and drug therapy, disseminated intravascular coagulation, and antibiotic-induced acute interstitial nephritis. Patients typically present with microscopic hematuria, urinary red blood cell casts, pyuria and modest proteinuria (nephrotic range in 25% of patients), and variable degrees of renal failure. Rheumatoid factor is present in 10 to 70%, and circulating immune complexes in 90%. Serum complement levels are usually depressed. Renal biopsy reveals mild focal proliferative glomerulonephritis with mesangial and capillary wall deposition of IgG and C3 by immunofluorescence microscopy and subendothelial, mesangial, and subepithelial electron-dense deposits by electron microscopy. Occasional patients develop diffuse necrotizing glomerulonephritis with crescent formation and present with nephritic syndrome or RPGN. Endocarditis-associated glomerulonephritis typically has a good prognosis and resolves with eradication of the underlying infection.

Immune-complex glomerulonephritis is a complication in 1 to 4% of patients with *infected ventriculoatrial shunts*. Nephritis can manifest weeks to years after shunt insertion and usually presents with microscopic hematuria. Nephrotic syndrome occurs in 30 to 50%. The usual renal pathology is a membranoproliferative pattern, although diffuse proliferation can also occur. Immunofluorescence reveals IgM and C3 in the capillary wall and mesangial area, while subendothelial deposits and mesangial interposition are seen by electron microscopy. Up to one-third of patients may have residual renal impairment despite removal of the infected shunt and resolution of the infection.

Suppurative infections such as intrathoracic and intraabdominal abscesses, osteomyelitis, and dental abscesses have been associated with glomerulonephritis. The usual presentation is hematuria, urinary red blood cell casts, proteinuria, and acute renal failure. Oliguria and hypertension are common. Pathologic renal lesions include

mesangial proliferative, membranoproliferative, and diffuse proliferative glomerulonephritis with crescents. Immunofluorescence reveals mesangial and capillary wall deposition predominantly of C3, although IgG and IgM may also be seen.

Nephrotic syndrome is a complication in 0.3% of patients with secondary *syphilis* and 8% of patients with congenital syphilis. The usual pathology is membranous glomerulopathy; however, mild mesangial and endocapillary proliferation can occur. IgG and IgM are evident in affected regions by immunofluorescence microscopy, and treponemal antigens have been identified in diseased glomeruli. C3 and C4 are typically depressed in congenital syphilis. The treatment consists of penicillin to eradicate the infection.

Leprosy most commonly causes AA amyloidosis; however, a syndrome resembling acute poststreptococcal glomerulonephritis has also been described.

PROTOZOAN AND PARASITIC INFECTIONS

Transient proteinuria (50% of patients) and nephrotic syndrome (<1% of patients) are complications of infection with *Plasmodium falciparum*. Membranoproliferative glomerulonephritis is the usual pathologic lesion and may respond to eradication of infection. *Plasmodium malariae* has been associated with diffuse or focal proliferative glomerulonephritis, membranous glomerulopathy, and minimal change disease. Eradication of the malarial infection does not consistently induce remission of the nephrotic syndrome. *Schistosoma mansoni* causes nephrotic syndrome in 5 to 10% of patients, and progression toESRD is common. The usual pathology isMPGN or mesangial proliferative glomerulonephritis, although membranous glomerulonephritis and amyloidosis are occasionally seen. *Filiariasis* can trigger membranous glomerulonephritis (*Loa loa*) and occasionally induces proliferative glomerulonephritis (*Onchocerca volvulus*). *Congenital toxoplasmosis* infection occasionally induces immune-complex glomerulonephritis characterized by mesangial and subendothelial immune deposits that contain *Toxoplasma* antigens. Membranous glomerulopathy and proliferative glomerulonephritis are occasional complications of *hydatid disease* and *trichinosis*, respectively.

GLOMERULAR LESIONS ASSOCIATED WITH NEOPLASIA

Glomerulopathies associated with neoplasia include membranous glomerulopathy, minimal change disease, focal segmental glomerulosclerosis, immune-complex glomerulonephritis, fibrillary/immunotactoid glomerulonephritis,LCDD, and amyloidosis. Mild proteinuria is common in patients with *solid tumors*, but overt glomerulonephritis is rare. Occasional patients with solid tumors of the lung, gastrointestinal tract, breast, kidney, and ovary develop full-blown nephrotic syndrome, usually due to a membranous glomerulopathy. Estimates of the incidence of occult malignancy in patients presenting with membranous glomerulopathy range from 0.1 to 10%. Most authorities agree that an extensive search for malignancy is not indicated, unless there are other suggestive clinical features. As many as 35% of patients with renal cell carcinoma have mesangial deposition of IgG and C3 visible on immunofluorescence; however, morphologic abnormalities are detected in only 50% of these patients, and clinically significant glomerulopathy is rare. Glomerular amyloidosis has also been described in association

with this tumor.

An array of glomerular disease has been reported in patients with lymphoproliferative malignancy. Nephrotic syndrome is a recognized complication of *Hodgkin's lymphoma*, with 70% of cases due to minimal change disease. The latter may occur concurrently with (40 to 45%), precede (10 to 15%), or follow (40 to 50%) diagnosis of the malignancy. It is postulated that a lymphokine or other mediator released by malignant T lymphocytes perturbs podocyte function and alters glomerular permeability in this setting. Nephrotic syndrome typically resolves with successful treatment and relapses with recurrence of disease. Less frequent associations with Hodgkin's lymphoma include focal segmental glomerulosclerosis, membranous glomerulopathy,MPGN, proliferative glomerulonephritis, and crescentic glomerulonephritis. Minimal change disease, membranous glomerulopathy, MPGN, and crescentic glomerulonephritis have also been reported in patients with *non-Hodgkin's lymphoma*. Glomerulopathy in the context of leukemia is rare. MPGN can complicate *chronic lymphatic leukemia* and related *B cell lymphomas*, particularly when associated with cryoglobulinemia. Other glomerular lesions associated with *paraproteinemia* include primary amyloid, LCDD, proliferative glomerulonephritis induced by cryoglobulinemia, and fibrillary/immunotactoid glomerulopathy. Here again, the renal lesion frequently improves or resolves with successful treatment of the underlying malignancy.

(Bibliography omitted in Palm version)

Back to Table of Contents

276. HEREDITARY TUBULAR DISORDERS - *John R. Asplin*, *Fredric L. Coe*

The hereditary renal tubular disorders and their morphologic and functional abnormalities, mode of inheritance, and associated abnormalities are summarized in Table 276-1. The individual disorders are discussed in detail below.

AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

ETIOLOGY AND PATHOLOGY

Autosomal dominant polycystic kidney disease (ADPKD) has a prevalence of 1:300 to 1:1000 and accounts for approximately 10% of end-stage renal disease (ESRD) in the United States. Some 90% of cases are inherited as an autosomal dominant trait, and approximately 10% are spontaneous mutations.

GENETIC CONSIDERATIONS

Three forms ofADPKDhave been identified. ADPKD-1 accounts for 90% of cases, and the gene has been localized to the short arm of chromosome 16. The gene for ADPKD-2 has been mapped to the long arm of chromosome 4. The protein products of the two genes form the polycystin complex, which may regulate cell-cell or cell-matrix interactions. A defect in either of these proteins interrupts the normal function of the polycystin complex, resulting in the same phenotype for two distinct genetic abnormalities. ADPKD-2 appears to have a later age of onset of symptoms and renal failure than ADPKD-1. A third form has been documented but has not been mapped to a gene at this point.

The kidneys are grossly enlarged, with multiple cysts studding the surface of the kidney. The cysts contain straw-colored fluid that may become hemorrhagic. The cysts are spherical, vary in size from a few millimeters to centimeters, and are distributed evenly throughout the cortex and medulla. Only 1 to 5% of nephrons will develop cysts. Cysts form when a "second hit" causes a somatic mutation in the normal allele of a tubule cell, leading to monoclonal proliferation of the tubular epithelium. The remaining renal parenchyma reveals varying degrees of tubular atrophy, interstitial fibrosis, and nephrosclerosis.

CLINICAL FEATURES

The disease may present at any age but most frequently causes symptoms in the third or fourth decade. Patients may develop chronic flank pain from the mass effect of the enlarged kidneys. Acute pain indicates infection, urinary tract obstruction by clot or stone, or sudden hemorrhage into a cyst. Gross and microscopic hematuria are common, and impaired renal concentrating ability frequently leads to nocturia. Nephrolithiasis occurs in 15 to 20% of patients, calcium oxalate and uric acid stones being most common. Low urine pH, low urine citrate, and urinary stasis from distortion of the collecting system by cysts all play a role in stone formation. Hypertension is found in 20 to 30% of children and up to 75% of adults. It is secondary to intrarenal ischemia from distortion of the renal architecture, leading to activation of the renin-angiotensin system. Patients with hypertension have a much more rapid progression toESRD.

Urinary tract infection is common and may involve the bladder or renal interstitium (pyelonephritis) or infect a cyst (pyocyst). Pyocysts can be difficult to diagnose but are more likely to be present if the patient has positive blood cultures, new renal pain, or failed to improve clinically after a standard course of antibiotic therapy.

Progressive decline in renal function is common, with approximately 50% of patients developingESRD by age 60. However, there is considerable variation in age of onset of renal failure, even within the same family. Hypertension, recurrent infections, male sex, and early age of diagnosis are related to early onset renal failure. Renal failure usually progresses slowly; if a sudden decrement in kidney function occurs, ureteral obstruction from stone, clot, or compression by a cyst are likely causes. Patients usually have high hematocrits for their level of renal function, as erythropoietin production is high. Fluid overload is uncommon because of a tendency for renal salt wasting.

Extrarenal manifestations of this disease are frequent and underscore the systemic nature of the defect. Hepatic cysts occur in 50 to 70% of patients. Cysts are generally asymptomatic, and liver function is normal, though women may develop massive hepatic cystic disease on occasion. Cyst formation has also been observed in the spleen, pancreas, and ovaries. Intracranial aneurysms are present in 5 to 10% of asymptomatic patients, with potential for permanent neurologic injury or death from subarachnoid hemorrhage. Screening of allADPKDpatients for aneurysms is not recommended, but patients with a family history of subarachnoid hemorrhage should be studied noninvasively with magnetic resonance imaging angiography. Colonic diverticular disease is the most common extrarenal abnormality, and patients are more likely to develop perforation than the general population with colonic diverticula. Mitral valve prolapse is found in 25% of patients, and the prevalence of aortic and tricuspid valve insufficiency is increased.

DIAGNOSIS

Ultrasound is the preferred technique for diagnosis of symptomatic patients and for screening asymptomatic family members. The ability to detect cysts increases with the subject's age: 80 to 90% ofADPKDpatients over the age of 20 will have detectable cysts, and almost 100% over the age of 30 will have cysts. At least three to five cysts in each kidney is the standard diagnostic criteria for ADPKD. Computed tomography (CT) scan may be more sensitive than ultrasound in detection of small cysts. Genetic linkage analysis is now available for diagnosis of ADPKD but is reserved for cases where radiographic imaging is negative and the need for definitive diagnosis critical, such as screening family members for potential kidney donation.

TREATMENT

The goals of treatment are to slow the rate of progression of renal disease and minimize symptoms. Hypertension and renal infection should be treated aggressively to maintain renal function. Converting enzyme inhibitors are effective antihypertensive agents, though patients should be closely monitored as some develop renal insufficiency and hyperkalemia. Urinary infection is treated in a standard manner unless a pyocyst is suspected, in which case antibiotics that penetrate cysts should be used, such as trimethoprim-sulfamethoxazole, ciprofloxacin, and chloramphenicol. Chronic pain from

cysts can be managed by cyst puncture and sclerosis with ethanol.

AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE

GENETIC CONSIDERATIONS

Autosomal recessive polycystic kidney disease (ARPKD) is a rare genetic disease that has an incidence between 1:10,000 and 1:40,000. The gene for ARPKD has been localized to chromosome 6. In the past, ARPKD was considered to be a family of disorders, categorized as neonatal, infantile, and childhood forms depending on the age of onset and the relative degree of involvement of the kidneys and liver. However, variable clinical presentations within siblings in the same family, as well as the localization of the disease to chromosome 6 in multiple families, support the premise that this is a single genetic disease with variable phenotypic presentation.

At birth the kidneys are enlarged with a smooth external surface. The distal tubules and collecting ducts are dilated into elongated cysts that are arranged in a radial fashion. As the patient ages, the cysts may become more spherical and the disease can be confused withADPKD. Interstitial fibrosis is also seen as renal function deteriorates. Liver involvement includes proliferation and dilation of small intrahepatic bile ducts as well as periportal fibrosis.

CLINICAL FEATURES

The majority of cases are diagnosed in the first year of life, presenting as bilateral abdominal masses. Death in the neonatal period is most commonly due to pulmonary hypoplasia. Hypertension and impaired urinary concentrating ability are common. The time course toESRD is variable, though many children maintain adequate kidney function for years. Older children present with complications secondary to congenital hepatic fibrosis and generally have less severe kidney disease. Hepatosplenomegaly, portal hypertension, and esophageal varices are frequent complications ofARPKD.

DIAGNOSIS

Ultrasound is the most common technique used to diagnoseARPKD, prenatally and in childhood. Ultrasound examination reveals enlarged kidneys with increased echogenicity. At times spherical cysts may be seen, potentially leading to an incorrect diagnosis ofADPKD. A thorough family history and imaging the kidneys of the parents aids in differentiation from other cystic diseases. The recent mapping of the gene should allow linkage studies to be used in diagnosis.

TREATMENT

Aggressive treatment of hypertension and urinary tract infection are the major goals of therapy in order to maintain native renal function as long as possible. Dialysis and transplant are appropriate when kidney failure occurs. Hepatic fibrosis may lead to life-threatening variceal hemorrhage, requiring sclerotherapy or portocaval shunting.

TUBEROUS SCLEROSIS

Patients with this multisystem disease most commonly present with skin lesions and benign tumors of the central nervous system (Chap. 370). Renal involvement is common; angiomyolipomas are the most frequent abnormality and are usually bilateral. Renal cysts may be present as well and can give an appearance similar to that ofADPKD. Histologically, the cysts are unique -- the cyst lining cells are large with an eosinophilic staining cytoplasm and may form hyperplastic nodules that can fill the cyst space.

GENETIC CONSIDERATIONS

One-third of cases are inherited as an autosomal dominant trait, the rest are due to sporadic mutations. Mutations of tumor-suppression genes have been identified on chromosomes 9 (*TSCI*) and 16 (*TSC2*). Mutations of *TSC2* account for the majority of cases and are more likely to be associated with mental retardation and polycystic kidneys. Tuberous sclerosis may be confused withADPKD if extrarenal manifestations are minimal.

VON HIPPEL-LINDAU DISEASE

This autosomal dominant disease is characterized by hemangioblastomas of the retina and the central nervous system (Chap. 370). Renal cysts occur in the majority of cases and are usually bilateral. The *VHL* gene is a tumor-suppressor gene and has been localized to chromosome 3. It is the same gene that is mutated in sporadic renal cell carcinoma, which may be found in up to 25% of patients with von Hippel-Lindau disease and is frequently multifocal. Yearly screening of adults using CT scans has been recommended in an attempt to diagnose renal cell cancers at an early stage.

MEDULLARY SPONGE KIDNEY

ETIOLOGY AND PATHOLOGY

Medullary sponge kidney (MSK) is a congenital disorder. Although some cases have apparent autosomal dominant inheritance, most are sporadic. It is found in 0.5 to 1% of all intravenous pyelograms. Males and females are affected equally. The pathologic lesion is cystic dilation of the inner medullary and papillary collecting ducts, with collecting diameters ranging from 1 to 5 mm. Bilateral renal involvement is present in 70% of cases, but not all papillae are equally affected. The dilated ducts are lined by cuboidal epithelium with areas of pseudostratified and stratified squamous epithelium. Calculi are frequently found in the dilated collecting ducts.

CLINICAL FEATURES

Patients generally present in the third or fourth decade with kidney stones, infection, or recurrent hematuria. The disease is most commonly diagnosed by intravenous pyelogram, which shows linear striations radiating into the renal papillae or small cystic collections of contrast in the dilated ducts (Fig. 276-1). Approximately 60% of patients withMSK have stones, and 12% of all stone formers will have MSK. Hypercalciuria occurs with the same frequency in MSK as it does in random stone formers. Papillary

nephrocalcinosis occurs more frequently in patients with MSK than in the random stone former. Proteinuria is minimal, if present at all, and renal function is normally preserved unless there is renal damage from recurrent infection or severe stone disease.

TREATMENT

Asymptomatic patients require no specific therapy except to maintain high fluid intake to reduce the risk of nephrolithiasis. If stones are present, standard laboratory evaluation should be done and metabolic abnormalities treated as in any stone former (Chap. 279). Infection should be treated aggressively, and instrumentation of the urinary tract should be minimized to avoid introducing infection.

JUVENILE NEPHRONOPHTHISIS/MEDULLARY CYSTIC DISEASE

ETIOLOGY AND PATHOLOGY

Juvenile nephronophthisis (JN) and medullary cystic disease (MCD) have similar pathologic findings but differ in inheritance pattern and age of onset.

GENETIC CONSIDERATIONS

JNis inherited as an autosomal recessive disease; linkage studies have shown 70% of the cases to map to a gene (*NPH1*) on the short arm of chromosome 2.MCD is an autosomal dominant disease. Linkage analysis has identified genes on chromosomes 1 and 16 as being associated with MCD. In both conditions, the kidneys tend to be small, with cysts throughout the medulla; the cortex and papilla rarely have cysts. The cysts originate in the collecting ducts, distal convoluted tubules, and loops of Henle and range in size from 1 to 10 mm. Sclerotic glomeruli, tubule atrophy, and interstitial fibrosis are frequent findings on biopsy.

CLINICAL FEATURES

Patients withJNpresent during childhood with symptoms of polyuria, growth retardation, anemia, and progressive renal insufficiency. Most patients developESRDprior to the age of 20; JN accounts for 2 to 10% of renal failure in children. Hepatic fibrosis and cerebellar ataxia has been reported in association with JN. JN with retinal degeneration is termed the *Senior-Loken syndrome*; it does not link to the *NPH1* gene at chromosome 2.MCDpresents in the third or fourth decade, though some cases may be diagnosed in the elderly population. Presenting symptoms in MCD are the same as in JN except for growth retardation. In addition, MCD does not have extrarenal abnormalities. Severe salt wasting can be seen, though this is usually a transient phase that resolves as the disease progresses to ESRD. Other features of tubule damage are often found, including hyperkalemia and hyperchloremic metabolic acidosis. Proteinuria is mild, and hematuria is rare.

DIAGNOSIS

The diagnosis is suggested by a family history of renal disease. The pattern of inheritance and age of onset aid in distinguishingJN/MCDfrom other inherited diseases.

Radiographic studies show small kidneys, loss of the corticomedullary junction, and multiple cysts in the medulla.CT scan is more sensitive than ultrasound in making the diagnosis. Open renal biopsy, including medullary tissue, may be required for diagnosis in some cases.

TREATMENT

Treatment is mainly supportive, as there is no specific therapy to prevent loss of renal function. Patients with salt wasting require a large oral intake of salt and water to maintain adequate extracellular volume. Alkali replacement and erythropoietin are required for acidosis and anemia, respectively. Renal transplantation has been performed in numerous patients, and the disease does not recur.

LIDDLE'S SYNDROME

Liddle's syndrome is a rare familial disease with a clinical presentation of hyperaldosteronism, consisting of hypertension, hypokalemia, and metabolic alkalosis. However, aldosterone levels are undetectable in these patients, and a nonaldosterone mineralocorticoid has not been isolated. Increased distal tubule sodium reabsorption, due to activating mutations in the amiloride-sensitive sodium channel, has been described in multiple families. Pharmacologic agents that block distal tubule sodium uptake, such as amiloride and triamterene, are effective in treating the hypertension and electrolyte abnormalities. As expected, spironolactone is ineffective, since the disease is not mediated via the aldosterone receptor.

BARTTER'S SYNDROME

CLINICAL FEATURES

Hypokalemia secondary to renal potassium wasting, metabolic alkalosis, and normal to low blood pressure are the clinical features of Bartter's syndrome. Three phenotypes of Bartter's syndrome have now been recognized. *Antenatal Bartter's syndrome* is characterized by polyhydramnios and premature delivery. During infancy, episodes of fever and dehydration are common and can lead to growth retardation. Nephrocalcinosis secondary to hypercalciuria is frequent. The infants also have a characteristic facies consisting of a triangular face with prominent eyes and ears. Prostaglandin E production is very high. Most cases of *classic Bartter's syndrome* present during childhood. Symptoms such as weakness and cramps are secondary to the hypokalemia. Polyuria and nocturia are common due to the hypokalemia-induced nephrogenic diabetes insipidus. Growth retardation may be seen. The *Gitelman's variant of Bartter's syndrome* presents during adolescence or adulthood and generally has a milder course than Bartter's syndrome. The dominant features are fatigue and weakness. It is distinguished from Bartter's syndrome by hypocalciuria, hypomagnesemia with hypermagnesuria, and normal prostaglandin production. All three forms are inherited as autosomal recessive traits. Although rarely required for diagnosis, renal biopsy reveals hyperplasia of the juxtaglomerular apparatus and prominence of medullary interstitial cells, with variable degrees of interstitial fibrosis, though these are not pathognomonic for the syndrome.

PATHOGENESIS

The pathogenesis of Bartter's syndrome has long been a matter of debate as the distinction of the primary disorder from the secondary phenomena induced by volume depletion and hypokalemia is difficult.

GENETIC CONSIDERATIONS

Recently, mutations in several renal tubule transport proteins have been shown to be responsible for the syndrome. In antenatal and classic Bartter's syndrome, impaired Cl^- reabsorption in the thick ascending limb of the loop of Henle is the underlying defect. Inadequate Cl^-reabsorption causes volume depletion and activates the renin-angiotensin system. Distal delivery of NaCl and water are high in the presence of high aldosterone, promoting secretion of K^+ and H^+ ions. Prostaglandin overproduction is mediated by volume depletion, hypokalemia, and high angiotensin II and kallikrein levels. Increased prostaglandin production contributes to the severity of disease by inducing resistance to the pressor effects of angiotensin II and reducing reabsorption in the thick ascending limb of the loop of Henle. Mutations in the bumetanide-sensitive Na:K:2Cl channel, the apical ATP-regulated K^+channel, and the basolateral Cl^- channel have been described in classic and antenatal Bartter's. All of these mutations would lead to a loss of Cl^-reabsorption in the loop of Henle. In Gitelman's syndrome, mutations have been found in the thiazide-sensitive NaCl transporter. The reduced Na^+reabsorption in the distal convoluted tubule leads to volume depletion and hypokalemia, though not as severe as would result from a lesion in the loop of Henle. Loss of activity of the thiazide-sensitive transporter increases tubule calcium reabsorption, leading to the classic finding of hypocalciuria in Gitelman's syndrome.

DIAGNOSIS

Hypokalemia, metabolic alkalosis, and normal to low blood pressure are the clinical findings characteristic of Bartter's syndrome. The differential diagnosis includes vomiting, surreptitious diuretic abuse, and magnesium deficiency. Chronic vomiting can be diagnosed by a low urine Cl^- concentration. Magnesium deficiency causes kaluresis and alkalosis, simulating Bartter's syndrome. Serum and urine magnesium will be low in such cases. Diuretic abuse produces metabolic abnormalities indistinguishable from Bartter's syndrome. Urine should be screened for diuretics multiple times before the diagnosis of Bartter's is made in a patient without a family history of the disorder.

TREATMENT

Dietary intake of sodium and potassium should be liberal. Potassium supplements are usually required. Magnesium supplements are needed in patients with Gitelman's syndrome. Spironolactone will reduce potassium wasting. Prostaglandin synthetase inhibitors are useful in patients with antenatal and classic Bartter's syndrome but are of no benefit in Gitelman's syndrome. Angiotensin-converting enzyme inhibitors may be beneficial in some patients.

CONGENITAL NEPHROGENIC DIABETES INSIPIDUS

GENETIC CONSIDERATIONS

This rare genetic disorder is most commonly inherited as an X-linked disease, with full expression in males and variable penetrance in females. Vasopressin acts through two receptors; type 1 receptors are located in the vasculature, while type 2 receptors are found in the collecting ducts of the kidney. In nephrogenic diabetes insipidus (NDI), only the actions requiring type 2 receptors are abnormal. Inactivating mutations of the type 2 vasopressin receptor, located on the long arm of the X chromosome, are responsible for the renal resistance to vasopressin. Less frequently, NDI may be inherited as an autosomal recessive trait, in which mutations in the gene for the water channels in collecting duct cells (aquaporin 2) lead to abnormal cell routing of aquaporin 2.

CLINICAL FEATURES

The clinical presentation is that of persistent polyuria, dehydration, and hypotonic urine in the presence of hypernatremia. Vasopressin levels are appropriately elevated in the hypertonic state, but renal response is lacking. The onset of the disorder is in infancy. The recurrent hypernatremia may lead to seizures or mental retardation. Once old enough to satisfy their thirst, children will be clinically stable though in a chronic state of polyuria and polydypsia. Renal function is normal, and radiographic studies of the urinary system reveal dilated ureters and bladder secondary to the chronically high urine flow. Since the most common form of the disease is X-linked, most patients are male. Heterozygous females generally have mild concentrating defects, though a few have phenotypic expression similar to males due to skewed X-chromosome inactivation. In the autosomal recessive form, males and females are affected equally.

TREATMENT

Treatment is aimed at maintaining adequate hydration. In the infant, low-solute feedings and high water intake are generally adequate. Addition of a thiazide diuretic reduces urine flow by inhibiting sodium reabsorption in the distal convoluted tubule. This lowers free water production and, by causing extracellular volume contraction, increases proximal salt and water reabsorption, reducing delivery to the distal nephron. Administration of vasopressin and its analogues has no role in the management of this disorder.

RENAL TUBULAR ACIDOSIS

Renal tubular acidosis (RTA) is a disorder of renal acidification out of proportion to the reduction in glomerular filtration rate. RTA is characterized by hyperchloremic metabolic acidosis with a normal serum anion gap [$Na^+ - (Cl^- + HCO_3^-)$]. There are multiple forms of RTA, depending on which aspects of renal acid handling have been affected. Defective bicarbonate reabsorption in the proximal tubule, suppressed renal ammoniagenesis, and inadequate distal tublule proton secretion are the abnormalities that produce RTA. Three types of RTA exist (Table 276-2). Types 1 and 2 may be inherited or acquired. Type 4 is acquired and is associated with either hypoaldosteronism or tubular hyporesponsiveness to mineralocorticoids. Type 3 was formerly used to define distal RTA with bicarbonate wasting in children; however, the bicarbonaturia resolves with age and is not truly part of a pathologic process. The term *type 3 RTA* is no longer used.

TYPE 1 (DISTAL)RTA

In this disorder the distal nephron does not lower urine pH normally, either because the collecting ducts permit excessive back-diffusion of hydrogen ions from lumen to blood or because there is inadequate transport of hydrogen ions. Excretion of titratable acid is low, as inadequate proton secretion prevents titration of urinary buffers such as phosphate. Urine ammonium excretion is inappropriately low for the level of acidosis, as the defect in acidification reduces the ion trapping required for ammonium excretion. Urinary concentration and potassium conservation also tend to be impaired.

Chronic acidosis lowers tubule reabsorption of calcium, causing renal hypercalciuria and mild secondary hyperparathyroidism. Buffering of bone by the daily metabolic acid load contributes to hypercalciuria. Urine citrate excretion is low, as acidosis and hypokalemia stimulate proximal tubule reabsorption of citrate. The hypercalciuria, alkaline urine, and low levels of urine citrate, which normally complexes about 40% of urine calcium, cause calcium phosphate stones and nephrocalcinosis. Growth in children is stunted because of rickets; this growth defect responds to amelioration of the acidosis with alkali. In the adult, osteomalacia occurs. In both children and adults, bone diseases may result, in part, from acidosis-induced loss of bone material and inadequate production of 1,25-dihydroxyvitamin D_3[1,25$(OH)_2D_3$]. Since the kidney does not conserve potassium or concentrate the urine normally, polyuria and hypokalemia occur. With the stress of an intercurrent illness, acidosis and hypokalemia can be life-threatening.

GENETIC CONSIDERATIONS

Type 1RTA can be familial, with autosomal dominant as the most common form of inheritance. X-linked, autosomal recessive, and sporadic cases have been reported. Mutations in the chloride-bicarbonate exchange gene (*AE1*) have been found in the autosomal dominant form. The cause of the autosomal recessive form is not known at this time. Other hereditary diseases that cause type 1 RTA include galactosemia, Ehler-Danlos syndrome, Fabry's disease,MSK, Wilson's disease, and hereditary elliptocytosis. The majority of cases of type 1 RTA are secondary to a systemic disorder such as Sjogren's syndrome, hypergammaglobulinemia, chronic active hepatitis, or lupus.

Diagnosis The diagnosis of type 1 RTA is suggested by a normal anion gap metabolic acidosis with a simultaneous urine pH greater than 5.5. Osteomalacia or rickets and calcium phosphate stones or nephrocalcinosis support the diagnosis, though they are not present in all cases. Bicarbonaturia is not present, which distinguishes this disorder from type 2 RTA. If acidosis is not severe and urine pH is equivocal, the oral ammonium chloride (NH_4Cl) loading test should be carried out: 0.1 g (1.9 mmol) NH_4Cl per kilogram of body weight is administered, and blood and urine pH are measured repeatedly over the next 6 h. Although systemic acidosis worsens, urine pH does not fall below 5.5. Urinary tract infection must not be present during this test because bacteria may possess urease, which hydrolyzes urea to ammonia and produces an alkaline urine.

Chronic diarrheal states cause normal anion gap acidosis and hypokalemia; urine pH may be>5.5 if ammonium production is very high. The urine anion gap ($Na^+ + K^+ - Cl^-$)

can be used to estimate renal ammonium production and distinguishRTA from gastrointestinal bicarbonate loss. Normally the urine anion gap is positive, as unmeasured anions exceed unmeasured cations. If urine ammonium levels are high, urine chloride concentration increases to balance the charge. Unmeasured cation (predominantly ammonium) now exceeds unmeasured anion, and the urine anion gap is negative. During metabolic acidosis, a negative urine anion gap suggests an extrarenal cause of acidosis, whereas a positive urine anion gap suggests RTA. The urine anion gap cannot be used if there are large amounts of unmeasured anions, such as bicarbonate or ketones, in the urine.

TREATMENT

Alkali supplements are the standard therapy. Enough alkali is prescribed to titrate the daily metabolic acid production, usually in the range of 0.5 to 2.0 mmol/kg body weight in four to six divided doses per day. Sodium bicarbonate and Shohl's solution (1 mmol sodium citrate and 1 mmol citric acid per mL) are common treatments. Potassium alkali salts can be used if hypokalemia is a persistent problem. Citrate requires less frequent dosing than bicarbonate salts as it is metabolized to bicarbonate after absorption. The dose of alkali should be raised until acidosis and hypercalciuria are both eliminated, and the patients should be followed by measurements of serum potassium, chloride, and CO_2content approximately twice yearly. Requirements for alkali usually rise during intercurrent illnesses but are usually below 4 mmol/kg body weight per day. The relatives of patients with idiopathic type 1RTAshould be screened for this disorder, as timely treatment can prevent growth retardation in children. Incomplete RTA secondary to idiopathic hypercalciuria is best treated using thiazide diuretics in conjunction with potassium citrate (Chap. 279).

TYPE 2 (PROXIMAL) RTA

Type 2 RTA usually occurs as part of a generalized disorder of proximal tubule function, presenting as hyperchloremic acidosis with other features of Fanconi syndrome. Bicarbonate reabsorption in the proximal tubule is defective. At normal concentrations of plasma bicarbonate, large amounts of bicarbonate are delivered to the distal tubule, overwhelming the absorptive capacity of the distal tubule and resulting in bicarbonaturia. As plasma bicarbonate levels fall, the lower filtered load of bicarbonate can be reabsorbed by the proximal tubule, resulting in normal distal delivery of bicarbonate. At this point the distal nephron can acidify the urine normally, resulting in normal excretion of daily metabolic acid production, albeit at a low serum bicarbonate level. Hypophosphatemia and low calcitriol levels are common and may lead to rickets or osteomalacia. Hypercalciuria occurs, but stone formation is unusual since urine citrate levels are normal or high because of reduced proximal tubule citrate reabsorption. Type 2 RTA may be inherited as autosomal dominant, autosomal recessive, or X-linked disorder. It may be acquired in association with other diseases (see "Fanconi Syndrome") or be secondary to drugs that inhibit carbonic anhydrase activity, such as acetazolamide.

Type 2RTA may be distinguished from type 1 RTA by the ability to normally acidify urine during spontaneous or ammonium chloride-induced acidosis. Correction of acidosis with bicarbonate will result in bicarbonaturia in type 2 RTA but not type 1 RTA. Fractional

excretion of bicarbonate is >15% at normal or near-normal serum bicarbonate levels. In distal RTA it is<10%. It is unusual for serum bicarbonate levels to fall below 15 mmol/L in proximal RTA. The urine anion gap will be positive, as ammonium excretion is normal to handle daily acid production but is not elevated as in nonrenal causes of acidosis.

TREATMENT

Children should be treated to prevent growth retardation. Alkali must be given in large amounts daily, 5 to 15 mmol/kg body weight per day, because bicarbonate is rapidly excreted in the urine. A thiazide diuretic can be used in conjunction with a low-salt diet to reduce the amount of bicarbonate required. Potassium supplementation is often required.

TYPE 4RTA

In type 4 RTA, also called *hyperkalemic distal RTA*, distal tubule secretion of both potassium and hydrogen ions is abnormal, resulting in hyperchloremic acidosis with hyperkalemia. Type 4 RTA is an acquired disorder; a moderate degree of renal insufficiency is present in the majority of patients. Patients with type 4 RTA can be differentiated from patients with type 1 since they have an acid urine (pH < 5.5) during periods of acidosis (Table 276-2) and hyperkalemia. They differ from type 2 patients by having a fractional excretion of bicarbonate<10% and a daily bicarbonate requirement of 1 to 3 mmol/kg body weight per day. Because potassium and hydrogen ion excretion are abnormal, such patients are considered to have generalized distal nephron dysfunction due to either insufficient aldosterone production or intrinsic renal disease causing aldosterone resistance. The resulting hyperkalemia reduces proximal tubule ammonia production, in addition to the inadequate proton secretion, leading to inadequate excretion of the daily metabolic proton load. These patients have an acid urine despite reduced proton secretion because there is inadequate ammonia to buffer protons in the distal tubule. If buffer delivery to the distal nephron is increased, urine pH will rise despite persistent acidosis.

Type 4RTA due to inadequate aldosterone production has multiple etiologies. Hyporeninemic hypoaldosteronism is the most common cause of type 4 RTA. Plasma levels of renin and aldosterone are subnormal, even during extracellular volume depletion, and the most common causes of this are diabetic nephropathy and chronic tubulointerstitial nephropathies. Nonsteroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors, trimethoprim, and heparin can reduce aldosterone production and produce a type 4 RTA. Drug-induced type 4 RTA is usually seen in patients with preexisting renal insufficiency. Reduced aldosterone production may be due to adrenal disease, either occurring as an isolated defect or as part of a more generalized adrenal disorder (Chap. 331). Renin levels are normal to high in adrenal disorders.

Patients with tubular resistance to aldosterone present with the same clinical features as those with hyporeninemic hypoaldosteronism. A tubulointerstitial process damages the distal tubule, restricting potassium and hydrogen ion excretion, despite adequate aldosterone levels. Obstructive uropathy and sickle cell disease are the most common causes of acquired tubular resistance to aldosterone. Hyporeninemic

hypoaldosteronism can be found in addition to tubular aldosterone resistance in many patients. Spironolactone, a competitive inhibitor of the aldosterone receptor, produces an aldosterone-resistant state. Amiloride and triamterene are diuretics that block sodium transport in the distal nephron, blunting the effect of aldosterone on the distal tubule.

TREATMENT

This is aimed mainly at reducing serum potassium, as acidosis will usually improve once the hyperkalemic block of ammonium production is removed. All patients should be placed on a low-potassium diet. Any drug that suppresses aldosterone production or blocks aldosterone effect should be discontinued. Mineralocorticoid supplementation with fludrocortisone, 0.1 to 0.2 mg/d, will improve hyperkalemia and acidosis; however, the patients who also have a partial tubule resistance to mineralocorticoid will require a higher dose. Mineralocorticoid replacement may not be appropriate for patients with hypertension or a history of heart failure. In such situations, a loop diuretic with a liberal sodium intake can usually promote adequate potassium excretion. Exchange resins will reduce potassium levels but are usually not tolerated well enough to be used for long-term treatment.

PSEUDOHYPOALDOSTERONISM

GENETIC CONSIDERATIONS

This rare inherited disorder is transmitted as either an autosomal dominant or recessive trait. The autosomal dominant form is caused by mutations in the mineralocorticoid receptor gene; the autosomal recessive disease is caused by inactivating mutations in the amiloride-sensitive epithelial sodium channel. The inability to respond to aldosterone leads to hyperkalemia, metabolic acidosis, salt wasting, and volume depletion, which present during childhood. Plasma renin and aldosterone levels are elevated. Treatment includes salt supplements, alkali, and potassium restriction.

VITAMIN D DISORDERS

X-LINKED HYPOPHOSPHATEMIC RICKETS (See also Chap. 342)

This disorder, also called *vitamin D-resistant rickets*, is an X-linked dominant disorder characterized by hypophosphatemia with renal phosphate wasting, rickets, and short stature. Hypophosphatemia is present soon after birth; rachitic bowing of the legs develops when the child begins to walk. Children have growth retardation, which is limited almost entirely to the lower extremities. Dentition is delayed, and skull abnormalities are common. Females generally have less severe disease than males. Presentation in adults ranges from disabling bone pain to no active symptoms, but generally some physical sign of childhood disease, such as short stature or bowed legs, is present. Overgrowth of bone at joints or sites of muscle attachment may reduce the mobility of the joint or cause nerve entrapment.

Hypophosphatemia secondary to reduced renal phosphate reabsorption is the hallmark of the disease. Intestinal phosphate absorption is low, worsening hypophosphatemia. Serum calcium levels are usually normal, with low intestinal absorption and renal

excretion of calcium. Serum alkaline phosphatase and osteocalcin levels are elevated. Parathyroid hormone levels are normal, as would be expected with normal serum calcium. $1,25(OH)_2D_3$ levels are usually normal, though in the setting of hypophosphatemia $1,25(OH)_2D_3$ levels should be elevated. Inadequate 1a-hydroxylase activity appears to play some role in the disease. Linkage analysis has localized the gene to the Xp22.1 region of the X chromosome. The gene has been identified and appears to be related to a family of endopeptidase genes.

TREATMENT

The goal of therapy is to raise serum phosphorous to normal or near-normal levels to improve bone mineralization. Oral neutral phosphate, 1 to 4 g/d in four to six doses, combined with calcitriol is an effective therapy that improves growth rate, reduces bone pain, and leads to radiographically evident improvement of the bone disease. Patients should be closely monitored during therapy as they may develop nephrocalcinosis and renal insufficiency.

VITAMIN D-DEPENDENT RICKETS TYPE I

GENETIC CONSIDERATIONS

This is an autosomal recessive disorder in which $1,25(OH)_2D_3$ levels are very low but 25-hydroxyvitamin D levels are normal. The disease is caused by inactivating mutations in the gene encoding the 1a-hydroxylase enzyme, leading to a clinical syndrome of vitamin D deficiency.

Symptoms usually appear before the age of 2, including rickets and growth retardation. Levels of serum calcium and phosphorous are low, but that of alkaline phosphatase is elevated. Intestinal calcium absorption and urinary calcium excretion are low. Parathyroid hormone is elevated in response to the hypocalcemia, resulting in increased urinary phosphate losses.

TREATMENT

Calcitriol (0.5 to 1 ug/d) leads to rapid correction of the biochemical abnormalities and resolution of the bone disease. Calcium and phosphorous supplementation are usually not required.

VITAMIN D-DEPENDENT RICKETS TYPE II (See also Chap. 342)

End-organ resistance to $1,25(OH)_2D_3$ is the pathogenesis of this disorder. Serum calcium and phosphate levels are low, secondary hyperparathyroidism is present, and $1,25(OH)_2D_3$ levels are elevated. Inheritance is usually autosomal recessive, though sporadic cases have been reported. Most patients present during childhood with rickets, though some have a milder form of disease not recognized until adulthood. Alopecia is common and tends to be associated with the more severe childhood form of the disease. Multiple defects have been detected in $1,25(OH)_2D_3$ receptor interaction, including absent hormone binding to the receptor, decreased receptor affinity, abnormal hormone-receptor localization, and abnormalities of the DNA-binding domain of the

receptor. Pharmacologic doses of calcitriol (5 to 30 ug/d) along with mineral supplementation will improve the biochemical disorders and bone disease, though some patients have no response to massive doses of calcitriol.

ONCOGENIC OSTEOMALACIA

This syndrome generally occurs in adults with highly vascular mesenchymal tumors. Patients present with bone pain and muscle weakness. Symptoms may be present for years before the correct diagnosis is made. Over 90% of the tumors are benign, and most are found in the extremities or maxillofacial region. Hypophosphatemia secondary to renal phosphate wasting and low levels of $1,25(OH)_2D_3$are the major biochemical abnormalities. Serum calcium and parathyroid hormone levels are normal. It appears the tumor produces a humoral agent that reduces proximal tubule phosphate reabsorption and 1a-hydroxylase activity. Removal of the tumor leads to rapid resolution of the disease.

X-LINKED RECESSIVE NEPHROLITHIASIS

This disorder presents as calcium nephrolithiasis in male children and progresses to nephrocalcinosis and renal failure. Low-molecular-weight proteinuria and hypercalciuria are also prominent features of the disease. Kidney biopsy reveals tubular atrophy, interstitial fibrosis, and medullary calcifications. The gene has been mapped to the short arm of the X chromosome and encodes a voltage-gated chloride channel (CLC-5). Dent's disease has been mapped to the same gene and has a similar presentation, except for an increased incidence of rickets.

ISOLATED HYPOURICEMIA (See also Chap. 353)

This disorder is generally inherited as an autosomal recessive trait. Most commonly there is deficient urate reabsorption in the proximal tubule, though some patients have been demonstrated to oversecrete urate. Serum uric acid is usually<120 umol/L (2 mg/dL) and hyperuricosuria is common, possibly due to decreased intestinal urate excretion. Hypouricemia is usually an incidental finding, as patients with this disorder are asymptomatic except for an increased risk of nephrolithiasis. Other disorders associated with hypouricemia include Fanconi syndrome, Wilson's disease, Hodgkin's disease, and Hartnup disease. No treatment is required except for high fluid intake to prevent kidney stones. Alkali and allopurinol may be used to prevent stones if fluids alone are not sufficient. Hypercalciuria has been associated with isolated hypouricemia in some families.

SELECTED DISORDERS OF AMINO ACID TRANSPORT

HARTNUP DISEASE

This disorder is characterized by reduced intestinal absorption and renal reabsorption of neutral amino acids. The defect involves an amino acid transporter on the brush border of the jejunum and the proximal tubule. Intestinal absorption of free amino acids is reduced, though the neutral amino acids can be absorbed when present in di- and tripeptides. Degradation of unabsorbed tryptophan by intestinal bacteria produces

indolic acids that are absorbed and subsequently excreted at high levels in the urine of these patients. The disorder is inherited as an autosomal recessive trait, affecting males and females equally. Widespread screening of newborns has estimated an incidence of 1 in 24,000 live births.

The majority of individuals with this disorder are asymptomatic. Approximately 10 to 20% present with clinical symptoms similar to those seen in pellagra, including a photosensitive erythematous scaly rash, intermittent cerebral ataxia, delirium, and diarrhea. Short stature is noted in some patients. The symptoms are thought to be due to deficiency in the essential amino acid tryptophan and resultant inadequate synthesis of nicotinamide. Though the inheritance of the disorder is Mendelian autosomal recessive, the development of symptomatic disease appears to be multifactorial. Diet, environment, and polygenic traits controlling plasma amino acid levels all contribute to development of symptoms.

Clinically affected patients can be differentiated from patients with pellagra by dietary history and the presence of aminoaciduria. Diagnosis is made by the characteristic finding of large amounts of neutral amino acids in the urine. It can easily be distinguished from generalized aminoaciduria by the normal excretion of proline. There are no other renal tubule defects as in Fanconi syndrome. Heterozygotes have normal urinary amino acid excretion.

TREATMENT

Symptomatic individuals should receive oral nicotinamide, 40 to 200 mg/d, and a high-protein diet to compensate for the poor amino acid absorption. Some patients who do not respond to nicotinamide may improve with tryptophan ethyl ester, which is lipid soluble and can be absorbed without an active transport system.

FANCONI SYNDROME

GENETIC CONSIDERATIONS

Fanconi syndrome is a generalized defect in proximal tubule transport involving amino acids, glucose, phosphate, uric acid, sodium, potassium, bicarbonate, and proteins. Idiopathic Fanconi syndrome may be inherited as an autosomal dominant, autosomal recessive, or X-linked trait. Sporadic cases are also seen. A variety of inherited systemic disorders are also associated with Fanconi syndrome including Wilson's disease, galactosemia, tyrosinemia, cystinosis, fructose intolerance, and Lowe's oculocerebral syndrome. The syndrome may be acquired in multiple myeloma, amyloid, and heavy metal toxicity.

The patients may present with a wide array of laboratory abnormalities including proximal renal tubular acidosis, glucosuria with a normal serum glucose, hypophosphatemia, hypouricemia, hypokalemia, generalized aminoaciduria, and low-molecular-weight proteinuria. Some patients do not have abnormalities in all proximal tubule transporters and may present with only a few of the laboratory findings. Rickets and osteomalacia are common findings secondary to the hypophosphatemia; production of calcitriol may also be abnormal. Metabolic acidosis also contributes to the

bone disease. Polyuria, salt wasting, and hypokalemia may be quite severe.

TREATMENT

Treatment includes phosphate supplements and calcitriol to heal the bone lesions, alkali for the acidosis, and liberal intake of salt and water. Alkali in the form of potassium salts may be particularly useful in the patient withRTA and hypokalemia. Aminoaciduria, glucosuria, hypouricemia, and low-molecular-weight proteinuria do not require treatment.

(Bibliography omitted in Palm version)

Back to Table of Contents

277. TUBULOINTERSTITIAL DISEASES OF THE KIDNEY - *Alan S. L. Yu, Barry M. Brenner*

Primary tubulointerstitial diseases of the kidney, as distinct from the disorders considered in Chaps. 274 and275, are characterized by histologic and functional abnormalities that involve the tubules and interstitium to a greater degree than the glomeruli and renal vasculature (Table 277-1). Secondary tubulointerstitial disease occurs as a consequence of progressive glomerular or vascular injury. These disorders can be further divided into acute and chronic forms. The chronic group may be due to sustained insults by a factor or factors that initially cause acute disease or to a slower, progressive, cumulative insult without an identifiable acute episode. Morphologically, acute forms of these disorders are characterized by interstitial edema, often associated with cortical and medullary infiltration by both mononuclear cells and polymorphonuclear leukocytes, and patchy areas of tubule cell necrosis. In more chronic forms, interstitial fibrosis predominates, inflammatory cells are typically mononuclear, and abnormalities of the tubules tend to be more widespread, as evidenced by atrophy, luminal dilatation, and thickening of tubule basement membranes. Because of the nonspecific nature of the histology, particularly in chronic tubulointerstitial diseases, biopsy specimens rarely provide a specific diagnosis. The urine sediment is also unlikely to be diagnostic, except in allergic forms of acute tubulointerstitial disease, in which eosinophils may predominate in the urinary sediment.

Defects in renal function often accompany these alterations of tubule and interstitial structure (Table 277-2). Proximal tubule dysfunction may be manifested as selective reabsorptive defects leading to hypokalemia, aminoaciduria, glycosuria, phosphaturia, uricosuria, or bicarbonaturia (proximal or type II renal tubular acidosis;Chap. 276). In combination, these defects constitute the *Fanconi syndrome*. Proteinuria, predominantly of low-molecular-weight proteins, is usually modest, rarely exceeding 2 g/d.

Defects in urinary acidification and concentrating ability often represent the most troublesome of the tubule dysfunctions encountered in patients with tubulointerstitial disease. Hyperchloremic metabolic acidosis often develops at a relatively early stage in the course. Patients with this finding generally elaborate urine of maximal acidity (pH £ 5.3). In such patients the defect in acid excretion is usually caused by a reduced capacity to generate and excrete ammonia due to the reduction in renal mass. Preferential damage to the collecting ducts, as in amyloidosis or chronic obstructive uropathy, may also predispose to distal or type I renal tubular acidosis (RTA), characterized by high urine pH (³5.5) during spontaneous or NH_4Cl-induced metabolic acidosis. Patients with tubulointerstitial diseases affecting predominantly medullary and papillary structures may also exhibit concentrating defects, with resultant nocturia and polyuria. Analgesic nephropathy and sickle cell disease are prototypes of this form of injury.

TOXINS

Although the kidneys constitute less than 1% of total body mass, they receive approximately 20% of the cardiac output, and 90% or more of renal blood flow is distributed to the renal cortex. Exposure of tubules and interstitium of the renal cortex to circulating toxins is therefore greater than for most other tissues. Transport processes in

renal tubules contribute further to the intrarenal accumulation of toxins, enhancing local concentrations of noxious agents. The urinary concentrating mechanism can also establish high levels of toxins within medullary and papillary portions of the kidney, predisposing these regions to chemical injury. Finally, the relatively acid pH of the fluid within most nephron segments may affect the ionization characteristics of potentially toxic compounds and thereby influence local concentration and solubility. Although these processes render the kidney vulnerable to toxic injury, the role of nephrotoxins in renal damage often goes unrecognized because the manifestations of such injury are usually nonspecific in nature and insidious in onset. Diagnosis largely depends on a history of exposure to a certain toxin. Particular attention should be paid to the occupational history, as well as to an assessment of exposure -- current and remote -- to drugs, especially antibiotics and analgesics, and to dietary supplements or herbal remedies. The recognition of a potential association between a patient's renal disease and exposure to a nephrotoxin is crucial, because, unlike many other forms of renal disease, progression of the functional and morphologic abnormalities associated with toxin-induced nephropathies may be prevented, and even reversed, by eliminating additional exposure.

EXOGENOUS TOXINS

Analgesic Nephropathy A distinct clinicopathologic syndrome has been described in heavy users of analgesic mixtures containing phenacetin in combination with aspirin, acetaminophen, or caffeine. These individuals have an approximately 20-fold increased risk of end-stage renal disease (ESRD). Analgesic nephropathy has been an important cause of chronic renal failure in Australia, Switzerland, Sweden, Belgium, and the southeastern United States.

Morphologically, analgesic nephropathy is characterized by papillary necrosis and tubulointerstitial inflammation. At an early stage, damage to the vascular supply of the inner medulla (vasa recta) leads to a local interstitial inflammatory reaction and, eventually, to papillary ischemia, necrosis, fibrosis, and calcification. The susceptibility of the renal papillae to damage by phenacetin is believed to be related to the establishment of a renal gradient for its acetaminophen metabolite, resulting in papillary tip concentrations tenfold higher than those in renal cortex. Hydration dissipates this gradient and may explain the protective effect of this maneuver in preventing phenacetin-induced papillary necrosis in animals. Aspirin in these analgesic compounds contributes to renal injury by uncoupling oxidative phosphorylation in renal mitochondria and by inhibiting the synthesis of renal prostaglandins, which are potent endogenous renal vasodilator hormones.

Analgesic nephropathy occurs some three to five times more commonly in women than in men. A direct relationship exists between the total amount of analgesic compounds ingested and the degree of renal impairment. The intake of 1.0 g phenacetin per day for 1 to 3 years or the total ingestion of 2 kg phenacetin in combination with other analgesics appears to represent minimum requirements for the development of analgesic nephropathy.

Whether single-ingredient analgesics other than phenacetin, when used alone, cause renal disease is controversial. Recent reports have suggested a two- to threefold

increase in the risk ofESRDamong regular users of acetaminophen, and perhaps nonsteroidal anti-inflammatory drugs (NSAIDs), but not among regular users of aspirin. Until conclusive evidence is available, physicians should consider screening regular users of acetaminophen and NSAIDs for evidence of renal disease.

In analgesic nephropathy, renal function usually declines gradually, in association with chronic necrosis of papillae and diffuse tubulointerstitial damage to the renal cortex. Occasionally, papillary necrosis may be associated with hematuria and even renal colic owing to obstruction of a ureter by necrotic tissue. More than half of patients with analgesic nephropathy have pyuria, which, if persistently associated with sterile urine, provides an important clue to the diagnosis. Nonetheless, active pyelonephritis may coexist in patients with analgesic nephropathy. Proteinuria, if present, is typically mild (< 1 g/d). Patients with analgesic nephropathy are usually unable to generate maximally concentrated urine, reflecting the underlying medullary and papillary damage. An acquired form of distalRTA may contribute to the development of *nephrocalcinosis*. The occurrence of anemia out of proportion to the degree of azotemia also may provide a clue to the diagnosis of analgesic nephropathy. When analgesic nephropathy has progressed to renal insufficiency, the kidneys usually appear bilaterally shrunken on intravenous pyelography, and the calyces are deformed. A "ring sign" on the pyelogram is pathognomonic of papillary necrosis and represents the radiolucent sloughed papilla surrounded by the radiodense contrast material in the calyx. Renal sonography may reveal papillary calcifications surrounding the central sinus complex in a "garland" pattern. Transitional cell carcinoma may develop in the urinary pelvis or ureters as a late complication of analgesic abuse.

TREATMENT

Every effort must be made to convince the patient who ingests excessive amounts of analgesic combinations to discontinue this hazardous practice. When renal damage is at an early stage, cessation of abuse usually arrests the progression of the nephrotoxic process; not infrequently, overall renal function improves with time. With continued abuse, however, progressive renal damage leads invariably to chronic renal failure.

Lead Nephropathy (See also Chap. 395) Lead intoxication may produce a chronic tubulointerstitial renal disease. Children who repeatedly ingest lead-based paints may develop kidney disease as adults. Significant occupational exposure may occur in a diverse variety of workplaces where lead-containing metals or paints are heated to high temperatures, such as battery factories, smelters, salvage yards, and firing ranges. Alcohol, illegally distilled in an apparatus constructed from automobile radiators (so-called moonshine), is another cause of lead poisoning. Environmental lead exposure, particularly in industrial regions, may be great enough to produce changes in renal function.

Tubule transport processes enhance the accumulation of lead within renal cells, particularly in the proximal convoluted tubule, leading to cell degeneration, mitochondrial swelling, and eosinophilic intranuclear inclusion bodies rich in lead. In addition to tubule degeneration and atrophy, lead nephropathy is associated with ischemic changes in the glomeruli, fibrosis of the adventitia of small renal arterioles, and focal areas of cortical scarring. Eventually, the kidneys become atrophic. Urinary excretion of lead, porphyrin

precursors such asd-aminolevulinic acid and coproporphyrin, and urobilinogen, may be increased. Patients with chronic lead nephropathy are characteristically *hyperuricemic*, a consequence of enhanced reabsorption of filtered urate. Acute gouty arthritis (so-called saturnine gout) develops in about 50% of patients with lead nephropathy, in striking contrast to other forms of chronic renal failure in which de novo gout is rare (Chap. 347). Hypertension is also a complication. Therefore, in any patient with slowly progressive renal failure, atrophic kidneys, gout, and hypertension, the diagnosis of lead intoxication should be considered. Features of acute lead intoxication (abdominal colic, anemia, peripheral neuropathy, and encephalopathy) are usually absent.

The diagnosis may be suspected by finding elevated serum levels of lead. However, because blood levels may not be elevated even in the presence of a toxic total-body burden of lead, the quantitation of lead excretion following infusion of the chelating agent calcium disodium edetate is a more reliable indicator of serious lead exposure. While urinary excretion of more than 0.6 mg/d of lead is generally considered to be indicative of overt or potential toxicity, recent evidence suggests that even lead burdens of 0.15 to 0.6 mg/d may cause progressive loss of renal function.

TREATMENT

Treatment includes removing the patient from the source of exposure and augmenting lead excretion with a chelating agent such as calcium disodium edetate.

Miscellaneous Nephrotoxins Use of *lithium salts* for bipolar disorder has been associated with polyuria and polydipsia caused by tubulointerstitial disease. There are only rare reports of chronic renal insufficiency attributable to this agent. Renal function should be followed in patients taking this drug, and caution should be exercised if lithium is employed in patients with underlying renal disease.

The immunosuppressant *cyclosporine* causes both acute and chronic renal injury. The acute injury and the use of cyclosporine in transplantation are discussed inChap. 272. The chronic injury results in an irreversible reduction in glomerular filtration rate (GFR), with mild proteinuria and arterial hypertension. Hyperkalemia is a relatively common complication and results in part from tubule resistance to aldosterone. Hypomagnesemia due to urinary magnesium wasting is less common but can cause hypocalcemia. The histologic changes in renal tissue include patchy interstitial fibrosis and tubular atrophy. In addition, the intrarenal vasculature often demonstrates hyalinosis, and focal segmental glomerular sclerosis can be present as well. Fibrosis may be the result of a cyclosporine-induced increase in renal collagen production. Vasoconstrictive mediators, such as angiotensin II, or vasoconstriction itself may also play a role in chronic cyclosporine toxicity. In patients receiving this drug for renal transplantation (Chap. 272), chronic rejection and recurrence of the primary disease may coincide with chronic cyclosporine injury, and on clinical grounds, distinction among these may be difficult. Although most patients experience stable, albeit reduced, renal function, progressive renal injury can occur without a progressive reduction in GFR. Dose reduction appears to mitigate cyclosporine-associated renal fibrosis but may increase the risk of rejection and graft loss. The optimal dosage of cyclosporine in renal transplantation remains controversial. Treatment of any associated arterial hypertension may lessen renal injury.

Many agents that commonly lead to acute renal failure are also capable of producing tubulointerstitial injury (Chap. 269). These include antibiotics (e.g., aminoglycosides, amphotericin B), radiographic contrast agents, various hydrocarbons (e.g., carbon tetrachloride), and heavy metals (e.g., mercury, cadmium, and bismuth).

METABOLIC TOXINS

Acute Uric Acid Nephropathy (See also Chap. 322) Acute overproduction of uric acid and extreme hyperuricemia often lead to a rapidly progressive renal insufficiency, so-called acute uric acid nephropathy. This tubulointerstitial disease is usually seen as part of the tumor lysis syndrome in patients given cytotoxic drugs for the treatment of lymphoproliferative or myeloproliferative disorders but may also occur in these patients before such treatment is begun. The pathologic changes are largely the result of deposition of uric acid crystals in the kidneys and their collecting systems, leading to partial or complete obstruction of collecting ducts, renal pelvis, or ureter. Since obstruction is often bilateral, patients typically show the clinical course of acute renal failure, characterized by oliguria and rapidly rising serum creatinine concentration. In the early phase uric acid crystals can be found in urine, usually in association with microscopic or gross hematuria. Hyperuricemia can also be a consequence of renal failure of any etiology. The finding of a urine uric acid-creatinine ratio greater than 1 mg/mg (0.7 mol/mol) distinguishes acute uric acid nephropathy from other causes of renal failure.

Prevention of hyperuricemia in patients at risk by treatment with allopurinol in doses of 200 to 800 mg/d prior to cytotoxic therapy reduces the danger of acute uric acid nephropathy. Once hyperuricemia develops, however, efforts should be directed to preventing deposition of uric acid within the urinary tract. Increasing urine volume with potent diuretics (furosemide or mannitol) effectively lowers intratubular uric acid concentrations, and alkalinization of the urine to pH 7 or greater with sodium bicarbonate and/or a carbonic anhydrase inhibitor (acetazolamide) enhances uric acid solubility. If these efforts, together with allopurinol therapy, are ineffective in preventing acute renal failure, dialysis should be instituted to lower the serum uric acid concentration as well as to treat the acute manifestations of uremia.

Gouty Nephropathy (See also Chap. 322) Patients with less severe but prolonged forms of hyperuricemia are predisposed to a more chronic tubulointerstitial disorder, often referred to as *gouty nephropathy*. The severity of renal involvement correlates with the duration and magnitude of the elevation of the serum uric acid concentration. Histologically, the distinctive feature of gouty nephropathy is the presence of crystalline deposits of uric acid and monosodium urate salts in kidney parenchyma. These deposits not only cause intrarenal obstruction but also incite an inflammatory response, leading to lymphocytic infiltration, foreign-body giant cell reaction, and eventual fibrosis, especially of medullary and papillary regions of the kidney. Bacteriuria and pyelonephritis occur in about one-fourth of cases, presumably as complications of intrarenal urinary stasis. Since patients with gout frequently suffer from hypertension and hyperlipidemia, degenerative changes of the renal arterioles may constitute a striking feature of the histologic abnormality, often out of proportion to other morphologic defects. Clinically, gouty nephropathy is an insidious cause of renal insufficiency. Early

in its course,GFR may be near normal, often despite focal morphologic changes in medullary and cortical interstitium, proteinuria, and diminished urinary concentrating ability. Whether reducing serum uric acid levels with allopurinol exerts a beneficial effect on the kidney remains to be demonstrated. Although such undesirable consequences of hyperuricemia as gout and uric acid stones respond well to allopurinol, use of this drug in asymptomatic hyperuricemia has not been shown to improve renal function consistently. On the other hand, uricosuric agents such as probenecid, which may increase uric acid stone production, clearly have no role in the treatment of renal disease associated with hyperuricemia.

Hypercalcemic Nephropathy (See also Chap. 341) Chronic hypercalcemia, as occurs in primary hyperparathyroidism, sarcoidosis, multiple myeloma, vitamin D intoxication, or metastatic bone disease, can cause tubulointerstitial damage and progressive renal insufficiency. The earliest lesion is a focal degenerative change in renal epithelia, primarily in collecting ducts, distal convoluted tubules, and loops of Henle. Tubule cell necrosis leads to nephron obstruction and stasis of intrarenal urine, favoring local precipitation of calcium salts and infection. Dilatation and atrophy of tubules eventually occur, as do interstitial fibrosis, mononuclear leukocyte infiltration, and interstitial calcium deposition (nephrocalcinosis). Calcium deposition also may occur in glomeruli and the walls of renal arterioles.

Clinically, the most striking defect is an inability to concentrate the urine maximally, resulting in polyuria and nocturia. Defective transport of NaCl in the ascending limb of Henle's loop is responsible, at least in part, for this concentrating defect. Additionally, reduced collecting duct responsiveness to vasopressin may contribute. Reductions in GFRand renal blood flow also occur, both in acute severe hypercalcemia and with prolonged hypercalcemia of lesser severity. DistalRTA and sodium and potassium wasting also have been described in these chronic states. Eventually, uncontrolled hypercalcemia leads to severe tubulointerstitial damage and overt renal failure. Abdominal x-rays may demonstrate nephrocalcinosis as well as nephrolithiasis, the latter due to the hypercalciuria that often accompanies hypercalcemia.

TREATMENT

This consists of reducing the serum calcium concentration toward normal and correcting the primary abnormality of calcium metabolism. The management of hypercalcemia is discussed in Chap. 341. Prognosis for recovery of renal function depends on the severity of the renal lesion at the time hypercalcemia is corrected. Renal dysfunction of acute hypercalcemia may be completely reversible. Gradual, progressive renal insufficiency related to chronic hypercalcemia, however, may not improve with correction of the calcium disorder. Nonetheless, every effort should be made to return serum calcium concentration to normal to minimize further loss of renal function.

Hypokalemic Nephropathy (See alsoChap. 49) Disturbances of renal structure and function occur commonly in patients with moderate to severe potassium depletion of at least several weeks' duration. Histologically, renal epithelial cells are often seen to contain numerous vacuoles, most marked in proximal tubules. Glomeruli are reduced in size and may become sclerotic. Whether prolonged or recurrent potassium deficiency results in irreversible tubulointerstitial fibrosis, scarring, and atrophy is unresolved. Loss

of urinary concentrating ability is the most commonly encountered functional defect and may be due to defective operation of the countercurrent multiplier system and elevated intrarenal prostaglandins. Nocturia, polyuria, and polydipsia are frequent symptoms. Urinalysis often reveals no abnormalities except for mild proteinuria. Serum creatinine and urea nitrogen concentrations usually remain within normal limits.

Miscellaneous Metabolic Toxins Urinary oxalate, derived from the metabolism of glycine and, to a variable extent, from ingested oxalate, may deposit as insoluble intratubular calcium oxalate crystals and result in chronic tubulointerstitial damage in patients with hereditary or acquired forms of *hyperoxaluria*. *Cystinosis* and *Fabry's disease* are other hereditary depositional disorders affecting the renal tubules and interstitium (Chap. 276).

RENAL PARENCHYMAL DISEASE ASSOCIATED WITH EXTRARENAL NEOPLASM

Except for the glomerulopathies associated with lymphomas and several solid tumors (Chap. 275), the renal manifestations of primary extrarenal neoplastic processes are confined mainly to the interstitium and tubules. Although metastatic renal involvement by solid tumors is unusual, the kidneys are often invaded by neoplastic cells in various lymphomas and leukemias and in multiple myeloma. In postmortem studies of patients with *lymphoma*, renal involvement is found in approximately half. The involvement may be focal, in the form of multiple discrete nodules, or diffuse, with lymphomatous infiltration throughout the renal parenchyma. Diffuse infiltration is seen most commonly in lymphomas other than Hodgkin's disease. There may be flank pain related to massive renal infiltration, and x-rays may show enlargement of one or both kidneys. Renal insufficiency occurs in a minority of cases, and overt uremia is rare. Treatment of the primary disease may improve renal function in these cases.

The kidneys are also commonly involved in various forms of *leukemia*. At postmortem examination, bilateral renal involvement is present in approximately 50% of cases. As with lymphoma, uremia is rarely, if ever, a consequence of leukemic infiltration of the kidneys. The kidneys can also be involved in leukemias because of the associated high incidence of hyperuricemia, hypercalcemia, and lysozymuria. The myelogenous leukemias, particularly of the monocytic type, may be complicated by tubule defects involving potassium and magnesium wasting.

PLASMA CELL DYSCRASIAS

Several glomerular and tubulointerstitial disorders may occur in association with plasma cell dyscrasias (Table 277-3;Chap. 113). Infiltration of the kidneys with myeloma cells is infrequent. When it occurs, the process is usually focal, so renal insufficiency from this cause is also uncommon. The more usual lesion is *myeloma kidney*, characterized histologically by atrophic tubules, many with eosinophilic intraluminal casts, and numerous multinucleated giant cells within tubule walls and in the interstitium. The frequent occurrence of myeloma kidney in patients with Bence Jones proteinuria has suggested a causal relation. Bence Jones proteins are thought to cause myeloma kidney through direct toxicity to renal tubule cells. In addition, Bence Jones proteins may precipitate within the distal nephron where the high concentrations of these proteins and the acid composition of the tubule fluid favor intraluminal cast formation and intrarenal

obstruction. Occasionally, acute renal failure occurs after intravenous pyelography in patients with multiple myeloma and is believed to result from the further precipitation of Bence Jones proteins induced by dehydration prior to radiographic study. Dehydration of the patient with myeloma in preparation for intravenous pyelography should therefore be avoided. Multiple myeloma may also affect the kidneys indirectly. Hypercalcemia or hyperuricemia may lead to the nephropathies described above. Proximal tubule disorders are also seen occasionally, including type II proximalRTA and the Fanconi syndrome.

AMYLOIDOSIS (See alsoChaps. 275 and319)

Glomerular pathology usually predominates and leads to heavy proteinuria and azotemia. However, tubule function may also be deranged, giving rise to a nephrogenic diabetes insipidus and to distal (type I)RTA. In several cases these functional abnormalities correlated with peritubular deposition of amyloid, particularly in areas surrounding vasa rectae, loops of Henle, and collecting ducts. Bilateral enlargement of the kidneys, especially in a patient with massive proteinuria and tubule dysfunction, should raise the possibility of amyloid renal disease.

IMMUNE DISORDERS

ALLERGIC INTERSTITIAL NEPHRITIS

An acute diffuse tubulointerstitial reaction may result from hypersensitivity to a number of drugs, including sulfonamides, many penicillins and cephalosporins, the fluoroquinolone antibiotics ciprofloxacin and norfloxacin, and the antituberculous drugs isoniazid and rifampin. Acute tubulointerstitial damage has also occurred after use of thiazide and loop diuretics, antiulcer medications (cimetidine, ranitidine, and omeprazole), andNSAIDs. Of note, the tubulointerstitial nephropathy that develops in some patients taking NSAIDs may be associated with nephrotic-range proteinuria and histologic evidence of either minimal change or membranous glomerulopathy. Grossly, the kidneys are usually enlarged. Histologically, the glomeruli appear normal. The principal pathologic abnormalities are in the interstitium of the kidney, which reveals pronounced edema and infiltration with polymorphonuclear leukocytes, lymphocytes, plasma cells, and, in some cases, large numbers of eosinophils. If the process is severe, tubule cell necrosis and regeneration may also be apparent. Immunofluorescence studies have either been unrevealing or demonstrated a linear pattern of immunoglobulin and complement deposition along tubule basement membranes. In a few cases of methicillin-induced acute tubulointerstitial disease, circulating anti-tubule basement membrane antibodies have also been found, suggesting that autoantibody formation may have been induced by the penicilloyl hapten of methicillin (by conjugation of hapten with tubule basement membrane proteins, thereby altering the native antigenicity of the basement membrane).

Most patients require several weeks of drug exposure before developing evidence of renal injury. Rare cases have occurred after only a few doses or after a year of more of use. Azotemia is usually present; a diagnostic triad of fever, skin rash, and peripheral blood eosinophilia is highly suggestive of acute tubulointerstitial nephritis but is often absent. Examination of the urine sediment reveals hematuria and often pyuria;

occasionally, eosinophils may be present. Proteinuria is usually mild to moderate, except in cases ofNSAID-induced tubulointerstitial nephritis with minimal change glomerulopathy. The clinical picture may be confused with acute glomerulonephritis, but when acute azotemia and hematuria are accompanied by eosinophilia, skin rash, and a history of drug exposure, a hypersensitivity reaction leading to acute tubulointerstitial nephritis should be regarded as the leading diagnostic possibility. Discontinuation of the drug usually results in complete reversal of the renal injury; rarely, renal damage may be irreversible. Glucocorticoids may accelerate renal recovery, but their value has not been definitively established.

SJOGREN'S SYNDROME (See alsoChap. 314)

When the kidneys are involved in this disorder, the predominant histologic findings are those of chronic tubulointerstitial disease. Interstitial infiltrates are composed primarily of lymphocytes, causing the histology of the renal parenchyma in these patients to resemble that of the salivary and lacrimal glands. Renal functional defects include diminished urinary concentrating ability and distal (type I)RTA. Urinalysis may show pyuria (predominantly lymphocyturia) and mild proteinuria.

TUBULOINTERSTITIAL ABNORMALITIES ASSOCIATED WITH GLOMERULONEPHRITIS

Primary glomerulopathies are often associated with damage to tubules and the interstitium. Occasionally, the primary disorder may affect glomeruli and tubules directly. For example, in more than half of patients with the nephropathy of systemic lupus erythematosus, deposits of immune complexes can be identified in tubule basement membranes, usually accompanied by an interstitial mononuclear inflammatory reaction. Similarly, in many patients with glomerulonephritis associated with anti-glomerular basement membrane antibody, the same antibody is reactive against tubule basement membranes as well. More frequently, tubulointerstitial damage is a secondary consequence of glomerular dysfunction. The extent of tubulointerstitial fibrosis correlates closely with the degree of renal impairment. Potential mechanisms by which glomerular disease might cause tublointerstitial injury include glomerular leak of plasma proteins toxic to epithelial cells, activation of tubule epithelial cells by glomerulus-derived cytokines, reduced peritubular blood flow leading to downstream tubulointerstitial ischemia, and hyperfunction of remnant tubules.

MISCELLANEOUS DISORDERS

VESICOURETERAL REFLUX (See also Chap. 281)

When the function of the ureterovesical junction is impaired, urine may reflux into the ureters due to the high intravesical pressure that develops during voiding. Clinically, reflux is often detected on the voiding and postvoiding films obtained during intravenous pyelography, although voiding cystourethrography may be required for definitive diagnosis. Bladder infection may ascend the urinary tract to the kidneys through incompetent ureterovesical sphincters. Not surprisingly, therefore, reflux is often discovered in patients with acute and/or chronic urinary tract infections. With more severe degrees of reflux, characterized by dilatation of ureters and renal pelves,

progressive renal damage often appears, and although active infection may also be present, uncertainty exists as to the necessity of infection in producing the scarred kidney of reflux nephropathy. Substantial proteinuria is often present, and glomerular lesions similar to those of idiopathic focal glomerulosclerosis (Chap. 274) are often found in addition to the changes of chronic tubulointerstitial disease. Surgical correction of reflux is usually necessary only with the more severe degrees of reflux since renal damage correlates with the extent of reflux. Obviously, if extensive glomerulosclerosis already exists, urologic repair may no longer be warranted.

RADIATION NEPHRITIS

Renal dysfunction can be expected to occur if 23 Gy (2300 rad) or more of x-ray irradiation is administered to both kidneys during a period of 5 weeks or less. Histologic examination of the kidneys reveals hyalinized glomeruli, atrophic tubules, extensive interstitial fibrosis, and hyalinization of the media of renal arterioles. Radiation-induced renal ischemia is believed to be the main pathogenic factor responsible for the tubulointerstitial damage, which may not become evident clinically for months after completion of radiation. The presentation of acute radiation nephritis includes rapidly progressive azotemia, moderate to malignant hypertension, anemia, and proteinuria that may reach the nephrotic range. More than 50% progress to chronic renal failure. A more insidious form is characterized by slower development of azotemia, anemia, and nephrotic syndrome. Malignant hypertension may follow unilateral renal irradiation and resolve with ipsilateral nephrectomy. Radiation nephritis has all but vanished because of heightened awareness of its pathogenesis by radiotherapists.

ACKNOWLEDGEMENT

Dr. Elliott Levy and Dr. Thomas H. Hostetter were co-authors of thischapterin the 14th edition and some of the material in thatchapteris carried forward to the present edition.

(Bibliography omitted in Palm version)

Back to Table of Contents

278. VASCULAR INJURY TO THE KIDNEY - *Kamal F. Badr*, *Barry M. Brenner*

Adequate delivery of blood to the glomerular capillary network is crucial for glomerular filtration and overall salt and water balance. Thus, in addition to the threat to the viability of renal tissue, vascular injury to the kidney may compromise the maintenance of body fluid volume and composition. Involvement of the renal vessels by atherosclerotic, hypertensive, embolic, inflammatory, and hematologic disorders is usually a manifestation of generalized vascular pathology. The morphologic and clinical responses to these insults and the unique renal vasculopathy associated with the toxemias of pregnancy are considered in this**chapter**.

THROMBOEMBOLIC DISEASES OF THE RENAL ARTERIES

Thrombosis of the major renal arteries or their branches is an important cause of deterioration of renal function, especially in the elderly. It is often difficult to diagnose and therefore requires a high index of suspicion. Thrombosis may occur as a result of intrinsic pathology in the renal vessels (posttraumatic, atherosclerotic, or inflammatory) or as a result of emboli originating in distant vessels, most commonly fat emboli, emboli originating in the left heart (mural thrombi following myocardial infarction, bacterial endocarditis, or aseptic vegetations), or "paradoxical" emboli passing from the right side of the circulation via a patent foramen ovale or atrial septal defect. Renal emboli are bilateral in 15 to 30% of cases.

The clinical presentation is variable, depending on the time course and the extent of the occlusive event. Acute thrombosis and infarction, such as follows embolization, may result in sudden onset of flank pain and tenderness, fever, hematuria, leukocytosis, nausea, and vomiting. If infarction occurs, renal enzymes may be elevated, namely aspartate aminotransferase (AST), lactic dehydrogenase (LDH; most reliable), and alkaline phosphatase, which rise and fall in the order listed. Urinary lactic dehydrogenase and alkaline phosphatase also may increase after infarction. Renal function deteriorates acutely, leading in bilateral thrombosis to acute oliguric renal failure. More gradual (i.e., atherosclerotic) occlusion of a single renal artery may go undetected. A spectrum of clinical presentations lies between these two extremes (Table 278-1). Hypertension usually follows renal infarction and results from renin release in the peri-infarction zone. Hypertension is usually transient but may be persistent. Diagnosis is established by renal arteriography.

TREATMENT

Management of *acute* renal arterial thrombosis includes surgical intervention, anticoagulant therapy, conservative and supportive therapy, and control of hypertension. The choice of treatment depends mainly on (1) the condition of the patient, in particular the patient's ability to withstand major surgery, and (2) the extent of renovascular occlusion and amount of renal mass at risk of infarction. In general, supportive care and anticoagulant therapy are indicated in unilateral disease. In bilateral thrombosis, medical and surgical therapies yield comparable results. Twenty-five percent of patients die during the acute episode, usually from extrarenal complications. In *chronic* ischemic renal disease, surgical revascularization is more likely to preserve and improve renal function and to control the hypertension (see below).

ATHEROEMBOLIC DISEASE OF THE RENAL ARTERIES

Atheroembolic disease typically results from multiple showers of cholesterol-containing microemboli dislodged from atheromatous plaques in large arteries. Such emboli occlude small (150- to 200-um diameter) vessels in the kidneys and in other organs (retina, brain, pancreas, muscles, skin, and extremities). Atheroembolic disease usually occurs in an elderly individual with atherosclerotic disease elsewhere and usually follows aortic surgery or renal or coronary arteriography. Spontaneous atheroembolic disease has also been reported. Manifestations include deterioration of renal function (sudden or gradual), mild proteinuria, microscopic hematuria, and leukocyturia. Urine volume may remain normal or fall to oliguric levels depending on severity. Renal ischemia can induce or exacerbate preexisting hypertension. In elderly patients with mild to moderate cholesterol embolization, the remaining nephrons may subsequently undergo injury, likely a result of hyperfiltration, which may lead to nephrotic-range proteinuria. Renal biopsy reveals "focal segmental sclerosis" (FGS). Renal function deteriorates at a slower rate in these individuals than in patients with more substantial embolic burdens, and they would be expected to benefit from angiotensin-converting enzyme (ACE) inhibitor therapy aimed at lowering intraglomerular pressures in remnant nephrons, even if their systemic blood pressure is in the "normal" range.

Antemortem diagnosis of atherosclerotic renal emboli is difficult. The demonstration of cholesterol emboli in the retina is helpful, but a firm diagnosis is established only by demonstration of cholesterol crystals in the smaller arteries and arterioles in renal biopsy or autopsy specimens. These also may be seen in asymptomatic skeletal muscle or skin. No specific treatment is available.

RENAL VEIN THROMBOSIS (RVT)

Thrombosis of one or both main renal veins occurs in a variety of settings (Table 278-2). The pathogenesis is not always clear, particularly when it occurs in so-called hypercoagulable states such as may develop in pregnant women, users of oral contraceptives, subjects with nephrotic syndrome, or dehydrated infants. Nephrotic syndrome accompanying membranous glomerulopathy and certain carcinomas seems to predispose to the development of RVT, which occurs in 10 to 50% of patients with these disorders. RVT may exacerbate preexisting proteinuria but is infrequently the cause of the nephrotic syndrome.

The clinical manifestations depend on the severity and abruptness of its occurrence. Acute cases occur typically in children and are characterized by sudden loss of renal function, often accompanied by fever, chills, lumbar tenderness (with kidney enlargement), leukocytosis, and hematuria. Hemorrhagic infarction and renal rupture may lead to hypovolemic shock. In young adultsRVT is usually suspected from an unexpected and relatively acute or subacute deterioration of renal function and/or exacerbation of proteinuria and hematuria in the appropriate clinical setting (underlying nephrotic syndrome, trauma, pregnancy, oral contraceptive use). In cases of gradual thrombosis, usually occurring in the elderly, the only manifestation may be recurrent pulmonary emboli or development of hypertension. A Fanconi-like syndrome and proximal renal tubular acidosis have been described.

The definitive diagnosis can only be established through selective renal venography with visualization of the occluding thrombus. Short of angiography, magnetic resonance imaging (MRI) often provides definitive evidence of thrombus.

TREATMENT

Treatment consists of anticoagulation, the main purpose of which is prevention of pulmonary embolization, although some authors have also claimed improvement in renal function and proteinuria. Encouraging reports have appeared concerning the use of streptokinase. Spontaneous recanalization with clinical improvement also has been observed. Anticoagulant therapy is more rewarding in the acute thrombosis seen in younger individuals. Nephrectomy is advocated in infants with life-threatening renal infarction. Thrombectomy is effective in some cases.

RENAL ARTERY STENOSIS/ISCHEMIC RENAL DISEASE

Stenosis of the main renal artery and/or its major branches accounts for 2 to 5% of hypertension (seeChap. 246). The common cause in the middle-aged and elderly is an atheromatous plaque at the origin of the renal artery. In a large unselected autopsy series, stenosis producing > 50% renal artery diameter reduction was found in 18% of those between 65 and 74 years of age and in 42% of those older than 75 years. Bilateral involvement was found in half of the affected cases in both age groups. Ischemic renal disease has emerged as an important cause of end-stage renal disease. It should be considered seriously in elderly individuals, particularly in those with evidence of atherosclerotic arterial disease elsewhere. In elderly patients with myocardial infarction or symptomatic peripheral vascular disease, the incidence of renal arterial stenosis can be up to 40%. In younger women, stenosis is due to intrinsic structural abnormalities of the arterial wall caused by a heterogeneous group of lesions termed *fibromuscular dysplasia*.

Renal artery stenosis should be suspected when hypertension develops in a previously normotensive individual over 50 years of age or in the young (under 30 years) with suggestive features: symptoms of vascular insufficiency to other organs, high-pitched epigastric bruit on physical examination, symptoms of hypokalemia secondary to hyperaldosteronism (muscle weakness, tetany, polyuria), and metabolic alkalosis. If renal arterial stenosis is suspected, the best initial screening test is a renal ultrasound, which may reveal unilateral renal hypotrophy (but normal cortical echogenicity). Absence of compensatory hypertrophy in the contralateral kidney should raise the suspicion of bilateral stensosis or superimposed intrinsic (structural) renal disease, most commonly hypertensive or diabetic nephropathy. A positive captopril test, which has a sensitivity and specificity of greater than 95%, constitutes an excellent follow-up procedure to assess the need for more invasive radiographic evaluation. The test relies on the exaggerated increase in plasma renin activity (PRA) after administration of captopril to patients with renovascular hypertension as compared with those with essential hypertension. It is considered positive when all the following criteria are satisfied: stimulated PRA of 12 (ug/L)/h, absolute increase in PRA of 10 (ug/L)/h or more, and increase in PRA of>150% [or 400% if baseline PRA is<3 (ug/L)/h]. BecauseACEinhibitors magnify the impairment in renal blood flow and glomerular

filtration rate (GFR) caused by functionally significant renal artery stenosis, use of these drugs in association with ^{99m}Tc-DTPA or ^{99m}MAG$_3$renography greatly enhances the predictive value of radionuclide renography (>90% sensitivity and specificity). Magnetic resonance angiography (MRA) has replaced previous modalities as the most sensitive (100%) and specific (95%) test for the diagnosis of renal arterial stenosis. The most definitive diagnostic procedure is bilateral arteriography with repeated bilateral renal vein and systemic renin determinations. If renal vein renin measurements from the two kidneys differ by a factor of 1.5:1 or more (higher value from the affected kidney) in a patient with radiographic unilateral renal artery stenosis, the chance of cure of hypertension by surgical reconstruction or angioplasty is almost 90%, particularly if the renal vein renin level from the unaffected kidney is equal to or less than systemic levels (suppressible). A ratio of less than 1.5:1, however, does not exclude the diagnosis of renovascular hypertension, particularly in the presence of bilateral disease.

TREATMENT

The aims of treatment are control of the blood pressure and restoration of perfusion to the ischemic kidney. In general, it is now firmly established that interventional therapy (i.e., surgery or angioplasty) is superior to medical therapy, which, while controlling blood pressure, does little to salvage renal mass lost to ischemic injury. Success rates with percutaneous transluminal angioplasty in young patients with fibromuscular dysplasia are 50% cure and improvement in blood pressure control in another 30%. Angioplasty is best suited for noncalcified, segmental short lesions and is also useful in some elderly patients who are poor surgical risks. About half of elderly individuals with reduced renal function as a result of renal arterial stenosis improve following angioplasty or surgery, even when preintervention arteriography shows little evidence of cortical perfusion. Despite the risks associated with surgery, long-term follow-up studies demonstrate an advantage of surgery over angioplasty both with regard to the incidence of restenosis and to the preservation or improvement inGFR. As with coronary angioplasty, stenting of renal arteries following balloon angioplasty is being used increasingly. Initial results are highly encouraging, with restenosis rates less than 15% at 6 months. Renal functional recovery or stabilization of renal function is seen in approximately 70% of patients. An illustrative example of renal artery stenting is shown inFig. 278-1.

Renal artery stenosis, particularly if atherosclerotic, is a progressive disease that may lead to gradual and silent loss of renal functional tissue (ischemic renal disease). Progression of ipsilateral atherosclerotic narrowing can be expected in nearly 50% of individuals, resulting in complete occlusion in about 10%. Thus, these patients need careful follow-up of initially nonclinically significant narrowing (<70%) for the possibility of further occlusion or the development of contralateral disease (30%). Compensatory contralateral hypertrophy may maintain renal function until affected by superimposed pathologic processes, at which time azotemia supervenes. Ischemic renal disease is now recognized as a significant cause of end-stage renal disease in patients over 50 years of age (approximately 15%). Even if angioplasty or surgery fail to return blood pressure to normal, these procedures usually render medical therapy easier.

HEMOLYTIC UREMIC SYNDROME (HUS) AND THROMBOTIC THROMBOCYTOPENIC PURPURA (TTP) (See alsoChap. 116)

HUS and TTP, consumptive coagulopathies characterized by microangiopathic hemolytic anemia and thrombocytopenia, have a particular predilection for the kidney and the central nervous system, the latter especially in TTP. The kidneys of patients with HUS or TTP often exhibit a "flea-bitten" appearance, the result of multiple cortical hemorrhagic infarcts. The major sites of pathology are the small renal arteries and afferent arterioles, which are nearly occluded as a result of marked intimal hyperplasia (particularly in TTP) and fibrin deposits in the subintimal regions. When the vasoocclusive process is extensive, bilateral cortical necrosis may occur. In addition, arteriolar microaneurysms, glomerular infarction, or nonspecific focal changes may be seen. In keeping with the focal nature of the vascular lesions, patchy areas of interstitial edema, tubular necrosis, and, eventually, fibrosis occur. By immunofluorescence staining, complement components and immunoglobulins may be demonstrated in the arterioles, and fibrinogen deposits are present in arteries, arterioles, and glomerular capillary loops.

Several mechanisms have been implicated in the etiology of the intravascular coagulopathy seen inHUS andTTP, including induction of a generalized Shwartzman phenomenon by microorganisms or endotoxin, genetic predisposition, and deficiency of platelet antiaggregatory substance(s) (e.g., prostacyclin). Some patients improve after exchange transfusion or plasmapheresis, suggesting accumulation of an as yet unidentified toxin.

Renal failure is common in bothHUS andTTP, usually manifested by azotemia, mild proteinuria, microscopic and/or gross hematuria, and cylindruria. Patients with HUS have more severe renal failure, often marked by oligoanuria and hypertension and commonly progressing to chronic renal failure. The prognosis in HUS is better in children than in adults. In TTP, the course of which may span days to months, renal failure is usually less severe.

TREATMENT

In the management ofTTP, high-dose glucocorticoids and plasma exchange often provide complete remission or cure. Plasma exchange should be initiated as early as possible, and the treatment cycles can be repeated if thrombocytopenia recurs. Splenectomy and antiplatelet therapy also have been used with varying degrees of success in patients with TTP. The success of plasma exchange in adultHUS is less well established than in TTP.

ARTERIOLAR NEPHROSCLEROSIS (See alsoChaps. 241 and246)

Whether hypertension is "essential" or of known etiology, persistent exposure of the renal circulation to elevated intraluminal pressures results in development of intrinsic lesions of the renal arterioles (hyaline arteriolosclerosis) that eventually lead to loss of function (nephrosclerosis). Nephrosclerosis is divided into two distinct entities: "benign" and "malignant" (or accelerated).

Benign Arteriolar Nephrosclerosis Benign arteriolar nephrosclerosis is seen in patients who are hypertensive for an extended period of time (blood pressure more than

150/90 mmHg) but whose hypertension has not progressed to a malignant form (described below). Such patients, usually in the older age group, are often discovered to be hypertensive on routine physical examination or as a result of nonspecific symptomatology (e.g., headaches, weakness, palpitations).

Kidney size is normal to reduced, with loss of cortical mass leading to a fine granularity. Although the larger arteries may show atherosclerotic changes, the characteristic pathology is in the afferent arterioles, which have thickened walls due to deposition of homogeneous eosinophilic material (hyaline arteriolosclerosis). This material is composed of plasma proteins and fats that have been deposited in the arteriolar wall due to injury to the endothelium, probably secondary to the elevated intraluminal hydraulic pressure. Narrowing of vascular lumina results, with consequent ischemic injury to glomeruli and tubules.

Nephrosclerosis accompanying long-standing systemic arterial hypertension is only one manifestation of a generalized process affecting the cardiovascular system. Physical examination, therefore, may reveal changes in retinal vessels (arteriolar narrowing and/or flame-shaped hemorrhages), cardiac hypertrophy, and possibly signs of congestive heart failure. Renal disease may manifest as a mild to moderate elevation of serum creatinine concentration, microscopic hematuria, and/or mild proteinuria. In general, clinical evaluation does not reveal significant renal abnormalities. More specialized examination may disclose elevated urinary albumin excretion, tapering and loss of caliber of intrarenal vessels on arteriography, and an exaggerated natriuresis in response to a fluid challenge. Patients with benign nephrosclerosis maintain a near-normalGFRdespite a reduction in renal blood flow.

Malignant Arteriolar Nephrosclerosis Patients with long-standing benign hypertension or patients not known to be hypertensive previously may develop malignant hypertension characterized by a sudden (accelerated) elevation of blood pressure (diastolic often above 130 mmHg) accompanied by papilledema, central nervous system manifestations, cardiac decompensation, and acute progressive deterioration of renal function. The absence of papilledema does not rule out the diagnosis in a patient with markedly elevated blood pressure and rapidly declining renal function. The kidneys are characterized by a flea-bitten appearance resulting from hemorrhages in surface capillaries. Histologically, two distinct vascular lesions can be seen. The first, affecting arterioles, is fibrinoid necrosis, i.e., infiltration of arteriolar walls with eosinophilic material including fibrin. There is thickening of vessel walls and, occasionally, an inflammatory infiltrate (necrotizing arteriolitis). The second lesion, involving the interlobular arteries, is a concentric hyperplastic proliferation of the cellular elements of the vascular wall with deposition of collagen to form a hyperplastic arteriolitis (onion-skin lesion). Fibrinoid necrosis occasionally extends into the glomeruli, which also may undergo proliferative changes or total necrosis. Most glomerular and tubular changes are secondary to ischemia and infarction. The sequence of events leading to the development of malignant hypertension is poorly defined. Two pathophysiologic alterations appear central in its initiation and/or perpetuation: (1) increased permeability of vessel walls to invasion by plasma components, particularly fibrin, which activates clotting mechanisms leading to a microangiopathic hemolytic anemia, thus perpetuating the vascular pathology; and (2) activation of the renin-angiotensin-aldosterone system at some point in the disease process, which

contributes to the acceleration and maintenance of blood pressure elevation and, in turn, to vascular injury.

Malignant hypertension is most likely to develop in a previously hypertensive individual, usually in the third or fourth decade of life. There is a higher incidence among men, particularly black men. The presenting symptoms are usually neurologic (dizziness, headache, blurring of vision, altered states of consciousness, and focal or generalized seizures). Cardiac decompensation and renal failure appear thereafter. Renal abnormalities include a rapid rise in serum creatinine, hematuria (at times macroscopic), proteinuria, and red and white blood cell casts in the sediment. Nephrotic syndrome may be present. Elevated plasma aldosterone levels cause hypokalemic metabolic alkalosis in the early phase. Uremic acidosis and hyperkalemia eventually obscure these early findings. Hematologic indices of microangiopathic hemolytic anemia (i.e., schistocytes) are often seen.

TREATMENT

Control of hypertension is the principal goal of therapy for both benign and malignant forms. The time of initiation of therapy, its effectiveness, and patient compliance are crucial factors in arresting the progression of benign nephrosclerosis. Untreated, most of these patients succumb to the extrarenal complications of hypertension. In contrast, malignant hypertension is a medical emergency; its natural course includes a death rate of 80 to 90% within 1 year of onset, almost always due to uremia. Supportive measures should be instituted to control the neurologic, cardiac, and other complications of acute renal failure, but the mainstay of therapy is prompt and aggressive reduction of blood pressure, which, if successful, can reverse all complications in the majority of patients. Presently, 5-year survival is 50%, and some patients have evidence of partial reversal of the vascular lesions and a return of renal function to near-normal levels.

SCLERODERMA (PROGRESSIVE SYSTEMIC SCLEROSIS) (See alsoChap. 313)

Renal vascular involvement in scleroderma is characterized by a distinctive lesion of the small arteries (diameters of 150 to 500 um) consisting of intimal proliferation, medial thinning, and increased collagen deposition in the adventitial layer. Fibrinoid changes in the walls of afferent arterioles and microinfarcts may occur. Glomerular changes are generally nonspecific and secondary to ischemic damage. Tubules are often atrophic. As part of a generalized increase in vasomotor tone, a vasospastic (Raynaud-like) phenomenon at the level of the renal vasculature contributes to the renal insufficiency. Reduction in renal blood flow is the major mechanism underlying the deterioration in kidney function, being present in 80% of patients, even in the absence of other clinical abnormalities. As vascular narrowing progresses, hypertension, azotemia, and proteinuria eventually develop. Plasma renin rises in response to sustained renal ischemia. The resulting hypertension causes further renal injury and may play a role in the ultimate destruction of nephrons. As more and more nephrons are lost to the combined insults of ischemia and hypertension, development of azotemia heralds a particularly grim prognosis. Proteinuria, usually mild, is a consequence of ischemic and hypertensive glomerular injury.

Although most patients with scleroderma present with extrarenal manifestations, renal

involvement is eventually manifested in half of patients followed for up to 20 years. Renal involvement can present in one of two ways, depending on whether malignant hypertension is superimposed on the renal pathology: (1) *Persistent urinary abnormalities* with or without hypertension tend to follow an indolent course with mild proteinuria, occasional casts, cellular elements in the urinary sediment, and a propensity for development of hypertension. Azotemia is absent initially, but when it develops, dialysis is required within 1 year. (2) *Scleroderma renal crisis* is a rapid deterioration in renal function, usually accompanied by malignant hypertension, oliguria, fluid retention, microangiopathic hemolytic anemia, and central nervous system involvement. It may occur in patients with previously undemonstrable or slowly progressive renal disease. Untreated, it leads to chronic renal failure within days to months.

The prognosis of scleroderma renal disease is generally poor, particularly following the onset of azotemia. Aggressive antihypertensive therapy may be effective in delaying the progression of renal failure. In scleroderma renal crisis, prompt treatment with beta blockers, minoxidil, and particularlyACEinhibitors may reverse acute renal failure. The effect of these interventions on renal function over the long term is uncertain.

SICKLE CELL NEPHROPATHY (See alsoChaps. 106 and275)

Sickle cell disease causes renal complications that arise mainly as a result of sickling of red blood cells in the microvasculature. The hypertonic and relatively hypoxic environment of the renal medulla, coupled with the slow blood flow in the vasa recta, favors the sickling of red blood cells, with resultant local infarction (papillary necrosis). Functional tubule defects in patients with sickle cell disease are likely the result of partial ischemic injury to the renal tubules.

In addition to the intrarenal microvascular pathology described above, young patients with sickle cell disease are characterized by renal hyperperfusion, glomerular hypertrophy, and hyperfiltration. Many of these individuals eventually develop a glomerulopathy leading to glomerular proteinuria (present in as many as 30%) and, in some, the nephrotic syndrome. In recent studies, the mechanisms underlying proteinuria in sickle cell nephropathy have been characterized as an early increase in pore radius, followed, as renal failure supervenes, with a reduction in pore number, but the onset of a dramatic loss of size-selectivity. Mild azotemia and hyperuricemia also can develop, but advanced renal failure and uremia are rare. Pathologic examination reveals the typical lesion of "hyperfiltration nephropathy," namely, focal segmental glomerular sclerosis. This finding has led to the suggestion that anemia-induced hyperfiltration in childhood is the principal cause of the adult glomerulopathy. Nephron loss secondary to ischemic injury also contributes to the development of azotemia in these patients.

In addition to the glomerulopathy described above, renal complications of sickle cell disease include the following: *Cortical infarcts* can cause loss of function, persistent hematuria, and perinephric hematomas. *Papillary infarcts*, demonstrated radiographically in 50% of patients with sickle trait, lead to an increased risk of bacterial infection in the scarred renal tissues and functional tubule abnormalities. Painless gross hematuria occurs with a higher frequency in sickle trait than in sickle cell disease and likely results from infarctive episodes in the renal medulla. *Functional tubule*

abnormalities such as nephrogenic diabetes insipidus result from marked reduction in vasa recta blood flow, combined with ischemic tubule injury. This concentrating defect places these patients at increased risk of dehydration and, hence, sickling crises. The concentrating defect also occurs in individuals with sickle trait. Other tubule defects involve potassium and hydrogen ion excretion, occasionally leading to hyperkalemic metabolic acidosis and a defect in uric acid excretion which, combined with increased purine synthesis in the bone marrow, results in hyperuricemia.

Management of sickle nephropathy is not separate from that of overall patient management (Chap. 106). In addition, however, the use ofACEinhibitors has been associated with improvement of the hyperfiltration glomerulopathy.

TOXEMIAS OF PREGNANCY (See also Chap. 7)

Renal function is "reset" at a higher level during normal pregnancy. Renal plasma flow and GFRboth increase by 30 to 50%. Therefore, serum creatinine levels above 70 umol/L (0.8 mg/dL) or blood urea nitrogen (BUN) levels above 4.6 mmol/L (13 mg/dL) are abnormal in pregnant women and should be investigated. Systolic and diastolic blood pressures decrease by an average of 10 to 15 mmHg below pregravid values. A diastolic pressure above 75 mmHg during the second trimester or above 85 mmHg during the third trimester is therefore abnormal. Vasodilation in the uterine, renal, and cutaneous beds, vasodilator prostaglandin release from the uteroplacental unit, and a decrease in arteriolar sensitivity to angiotensin II all play a role in the decline of blood pressure during pregnancy.

Preeclampsia-Eclampsia The toxemia syndrome, usually occurring in the third trimester of primigravidas, includes hypertension, proteinuria, edema, consumptive coagulopathy, sodium retention, hyperreflexia (preeclampsia), and, if uncontrolled, convulsions (eclampsia). In pure preeclampsia (i.e., not superimposed on previously existing hypertensive or renal disease), the primary sites of pathology are the glomerular endothelial cells. These cells show marked swelling due to an increase in cytoplasmic volume with vacuolization (endotheliosis) and encroach on the vascular lumen, rendering the enlarged glomeruli ischemic. The glomerular basement membrane and the extraglomerular blood vessels are intact. The pathogenesis is unknown. Coagulation abnormalities, hormonal factors, uteroplacental ischemia, and immune mechanisms have all been implicated. Increased microvascular reactivity may be a result of endothelial cell damage, which, in turn, alters the balance of endothelium-derived vasodilator/vasoconstrictor autacoids. Recent evidence suggests that preeclampsia may be characterized by selective dysregulation of vascular cell adhesion molecule-1 (VCAM-1) (but not other leukocyte adhesion molecules). This abnormality is not present in non-proteinuric gestational hypertension, and subsides post-partum. Induction of VCAM-1 expression in preeclampsia may contribute to leukocyte-mediated tissue injury in this condition or may reflect perturbation of other, previously unrecognized functions of this molecule in pregnancy. Despite sodium retention, intravascular volume is contracted as compared with pregravid values. An increased sensitivity to angiotensin II is the basis for the "roll-over test" (an increase in diastolic blood pressure of 20 mmHg or more on changing the patient's position from lateral recumbent to supine, presumably due to alterations in circulating angiotensin levels). In the supine position, the reduction in venous return due to compression by the

gravid uterus increases circulating levels of angiotensin II. This increase results in a hypertensive response in preeclamptic patients, who are hyperresponsive to angiotensin II, but not in normal women, in whom pregnancy leads to a relative resistance to the pressor effects of this hormone.

A diagnosis of preeclampsia-related hypertension can be made when repeated measurements over a 4- to 6-h period show a blood pressure of 140/85 mmHg or more. The rise in blood pressure tends to be more severe at night. When preeclampsia occurs in a previously hypertensive patient, a rapid acceleration of the blood pressure elevation is accompanied by an increase in proteinuria, oliguria, edema, and coagulopathy. This is a life-threatening syndrome and tends to recur with future pregnancies. In addition to proteinuria, which correlates with the severity of the renal lesion,GFR and renal plasma flow are depressed. In view of the preexisting high levels, however, GFR in preeclamptic women often remains above nonpregnant levels. Uric acid clearance also falls, resulting in hyperuricemia. In the postpartum period, these patients are particularly susceptible to the development of "postpartum renal failure," which is thought to be a form of adultHUS.

TREATMENT

Treatment consists of bed rest in a quiet environment and control of neurologic manifestations and blood pressure, the former with magnesium sulfate and the latter usually with vasodilators such as hydralazine and methyldopa. Diuretics are avoided. The ultimate "treatment" is delivery, which should be induced if fetal maturity is adequate or if life-threatening coagulopathy or renal failure occur. The long-term prognosis is generally favorable.

Development of acute renal failure/preeclampsia in a pregnant woman should alert the physician to potential preexisting renal disease and/or hypertension. The latter is particularly likely if systolic blood pressure is greater than 200 mmHg. Hypertension and preexisting proteinuria tend to worsen in 50% of women during pregnancy. In addition, these abnormalities may be unmasked during pregnancy as the first manifestations of an underlying glomerulopathy. Conversely, patients with established underlying renal disease should be followed closely during pregnancy, with monthly measurements of 24-h urinary protein excretion andGFR. Sudden deterioration should raise suspicion of superimposed preeclampsia. There is no convincing evidence that pregnancy has an adverse effect on the long-term outcome of immunologic glomerular diseases or diabetic nephropathy. In all situations, control of blood pressure should be the primary therapeutic goal in view of its established beneficial effects on the progression of renal injury.

Bilateral Cortical Necrosis Acute bilateral cortical necrosis is associated with septic abortions, abruptio placentae, and preeclampsia. Coagulation in cortical vessels and arterioles leads to renal tissue necrosis. Anuria and renal failure ensue and may be irreversible. In other cases, renal function returns partially, but on long-term follow-up most patients slowly progress to uremia.

(Bibliography omitted in Palm version)

Back to Table of Contents

279. NEPHROLITHIASIS - *John R. Asplin*, *Fredric L. Coe*, *Murray J. Favus*

TYPES OF STONES

Calcium salts, uric acid, cystine, and struvite ($MgNH_4PO_4$) are the basic constituents of most kidney stones in the western hemisphere. Calcium oxalate and calcium phosphate stones make up 75 to 85% of the total (Table 279-1) and may be admixed in the same stone. Calcium phosphate in stones is usually hydroxyapatite [$Ca_5(PO_4)_3OH$] or, less commonly, brushite ($CaHPO_4 \times H_2O$).

Calcium stones are more common in men; the average age of onset is the third decade. Approximately 60% of people who form a single calcium stone eventually form another within the next 10 years. The average rate of new stone formation in patients who have had a previous stone is about one stone every 2 or 3 years. Calcium stone disease is frequently familial.

In the urine, calcium oxalate monohydrate crystals (whewellite) usually grow as biconcave ovals that resemble red blood cells in shape and size but may occur in a larger, "dumbbell" form. In polarized light the crystals appear bright against a dark background, with an intensity that is dependent on orientation, a property known as *birefringence.* Calcium oxalate dihydrate crystals (weddellite) are bipyramidal. Apatite crystals do not exhibit birefringence and appear amorphous because the actual crystals are too small to be resolved by light microscopy.

Uric acid stones (Table 279-1) are radiolucent and are also more common in men. Half of patients with uric acid stones have gout; uric acid lithiasis is usually familial whether or not gout is present. In urine, uric acid crystals are red-orange in color because they absorb the pigment uricine. Anhydrous uric acid produces small crystals that appear amorphous by light microscopy. They are indistinguishable from apatite crystals, except for their birefringence. Uric acid dihydrate tends to form teardrop-shaped crystals as well as flat, rhomboid plates; both are strongly birefringent. Uric acid gravel appears like red dust, and the stones are also orange or red on some occasions. *Cystine stones* are uncommon (Table 279-1), lemon yellow, and sparkle; radiopacity is due to the sulfur content. Cystine crystals appear in the urine as flat, hexagonal plates.

Struvite stones are common (Table 279-1) and potentially dangerous. These stones occur mainly in women or patients who require chronic bladder catheterization and result from urinary tract infection with urease-producing bacteria, usually *Proteus* species. The stones can grow to a large size and fill the renal pelvis and calyces to produce a "staghorn" appearance. They are radiopaque and have a variable internal density. In urine, struvite crystals are rectangular prisms said to resemble coffin lids.

MANIFESTATIONS OF STONES

As stones grow on the surfaces of the renal papillae or within the collecting system, they need not produce symptoms. Asymptomatic stones may be discovered during the course of radiographic studies undertaken for unrelated reasons. Stones rank, along with benign and malignant neoplasms, renal cysts, and genitourinary tuberculosis, among the common causes of isolated hematuria. Much of the time, however, stones

break loose and enter the ureter or occlude the ureteropelvic junction, causing pain and obstruction.

STONE PASSAGE

A stone can traverse the ureter without symptoms, but passage usually produces pain and bleeding. The pain begins gradually, usually in the flank, but increases over the next 20 to 60 min to become so severe that narcotic drugs may be needed for its control. The pain may remain in the flank or spread downward and anteriorly toward the ipsilateral loin, testis, or vulva. Pain that migrates downward indicates that the stone has passed to the lower third of the ureter, but if the pain does not migrate, the position of the stone cannot be predicted. A stone in the portion of the ureter within the bladder wall causes frequency, urgency, and dysuria that may be confused with urinary tract infection. The vast majority of ureteral stones less than 0.5 cm in diameter will pass spontaneously.

It has been standard practice to diagnose acute renal colic by intravenous pyelography; however, helical computed tomography (CT) scan without radiocontrast enhancement is now the preferred procedure. CT has the advantage of detecting uric acid stones in addition to the traditional radioopaque stones, and CT does not expose the patient to the risk of radio-contrast agents.

OTHER SYNDROMES

Staghorn Calculi Struvite, cystine, and uric acid stones often grow too large to enter the ureter. They gradually fill the renal pelvis and may extend outward through the infundibula to the calyces themselves.

Nephrocalcinosis Calcium stones grow on the papillae. Most break loose and cause colic, but they may remain in place so that multiple papillary calcifications are found by x-ray, a condition termed *nephrocalcinosis*. Papillary nephrocalcinosis is common in hereditary distal renal tubular acidosis (RTA) and in other types of severe hypercalciuria. In medullary sponge kidney disease (Chap. 276) calcification may occur in dilated distal collecting ducts.

Sludge Sufficient uric acid or cystine in the urine may plug both ureters with precipitate. Calcium oxalate crystals do not do this because less than 100 mg oxalate usually is excreted daily in the urine even in severe hyperoxaluric states, compared with 1000 mg uric acid in patients with hyperuricosuria and 400 to 800 mg cystine in patients with cystinuria. Calcium phosphate crystals can render the urine milky but do not plug the urinary tract.

INFECTION

Although urinary tract infection is not a direct consequence of stone disease, it can occur after instrumentation and surgery of the urinary tract, which are frequent in the treatment of stone disease. Stone disease and urinary tract infection can enhance their respective seriousness and interfere with treatment. Obstruction of an infected kidney by a stone may lead to sepsis and extensive damage of renal tissue, since it converts

the urinary tract proximal to the obstruction into a closed, or partially closed, space that can become an abscess. Stones may harbor bacteria in the stone matrix, leading to recurrent urinary tract infection. On the other hand, infection due to bacteria that possess the enzyme urease can cause stones composed of struvite.

ACTIVITY OF STONE DISEASE

Active disease means that new stones are forming or that preformed stones are growing. Sequential radiographs of the renal areas are needed to document the growth or appearance of new stones and to ensure that passed stones are actually newly formed, not preexistent ones.

PATHOGENESIS OF STONES

Urinary stones usually arise because of the breakdown of a delicate balance. The kidneys must conserve water, but they must excrete materials that have a low solubility. These two opposing requirements must be balanced during adaptation to diet, climate, and activity. The problem is mitigated to some extent by the fact that urine contains substances that inhibit crystallization of calcium salts and others that bind calcium in soluble complexes. These protective mechanisms are less than perfect. When the urine becomes supersaturated with insoluble materials, because excretion rates are excessive and/or because water conservation is extreme, crystals form and may grow and aggregate to form a stone.

SUPERSATURATION

In a solution in equilibrium with crystals of calcium oxalate, the product of the chemical activities of the calcium and oxalate ions in the solution is termed the *equilibrium solubility product*. If crystals are removed, and if either calcium or oxalate ions are added to the solution, the activity product increases, but the solution may remain clear; no new crystals form. Such a solution is *metastably supersaturated.* If new calcium oxalate seed crystals are now added, they will grow in size. Ultimately, the activity product reaches a critical value at which a solid phase begins to develop spontaneously. This value is called the *upper limit of metastability*, or the *formation product*. Stone growth in the urinary tract requires a urine that, on average, is above the equilibrium solubility product. Excessive supersaturation is common in stone formation.

Calcium, oxalate, and phosphate form many stable soluble complexes among themselves and with other substances in urine, such as citrate. As a result, their free ion activities are below their chemical concentrations and can be measured only by indirect techniques. Reduction in ligands such as citrate can increase ion activity without changing total urinary calcium. Urine supersaturation can be increased by dehydration or by overexcretion of calcium, oxalate, phosphate, cystine, or uric acid. Urine pH is also important; phosphate and uric acid are weak acids that dissociate readily over the physiologic range of urine pH. Alkaline urine contains more dibasic phosphate, favoring deposits of brushite, and apatite. Below a urine pH of 5.5, uric acid crystals (p*K* 5.47) predominate, whereas phosphate crystals are rare. The solubility of calcium oxalate, on the other hand, is not influenced by changes in urine pH. Measurements of supersaturation in a pooled 24-h urine sample probably underestimate the risk of

precipitation. Transient dehydration, variation of urine pH, and postprandial bursts of overexcretion may cause values considerably above average.

NUCLEATION

Homogeneous Nucleation In urine that is supersaturated with respect to calcium oxalate, these two ions form clusters. Most small clusters eventually disperse because the internal forces that hold them together are too weak to overcome the random tendency of ions to move away. Clusters of over 100 ions can remain stable because attractive forces balance surface losses. Once they are stable, nuclei can grow at levels of supersaturation below that needed for their creation. The formation product marks the point at which stable nuclei become frequent enough to create a permanent solid phase.

Heterogeneous Nucleation If a supersaturated urine is seeded with preformed nuclei of a crystal that is similar in structure to calcium oxalate, calcium and oxalate ions in solution will bind to the crystal's surface as they would on a seed crystal of calcium oxalate itself. The seeding of a supersaturated solution by foreign nuclei is called *heterogeneous nucleation*. Cell debris, calcifications on the renal papillae, as well as other urinary crystals, can serve as heterogeneous nuclei that permit calcium oxalate stones to form, even though urine calcium oxalate supersaturation never exceeds the metastable limit for homogenous nucleation.

INHIBITORS OF CRYSTAL FORMATION

Stable nuclei must grow and aggregate to produce a stone of clinical significance. Urine contains potent inhibitors of nucleation, growth, and aggregation for calcium oxalate and calcium phosphate but not for uric acid, cystine, or struvite. Inorganic pyrophosphate is a potent inhibitor that appears to affect calcium phosphate more than calcium oxalate crystals. Citrate inhibits crystal growth and nucleation, though most of the stone inhibitory activity of citrate is due to lowering urine supersaturation via complexation of calcium. Other urine components such as glycoproteins inhibit all three processes of calcium oxalate stone formation. Slowing of crystal growth increases the apparent upper limit of metastability because the critical growth of ion clusters into stable nuclei is hindered. As a consequence of the presence of these inhibitors, crystal growth in urine is slow compared with growth in simple salt solutions, and the upper limit of metastability is higher.

EVALUATION AND TREATMENT OF PATIENTS WITH NEPHROLITHIASIS

Most patients with nephrolithiasis have remediable metabolic disorders that cause stones and can be detected by chemical analyses of serum and urine. Adults with recurrent kidney stones and children with even a single kidney stone should be evaluated. A practical outpatient evaluation consists of two or three 24-h urine collections, each with a corresponding blood sample; measurements of serum and urine calcium, uric acid, electrolytes and creatinine, urine pH, volume, oxalate, and citrate should be made. Since stone risks vary with diet, activity, and environment, at least one urine collection should be made on a weekend when the patient is at home and another on a work day. When possible, the composition of kidney stones should be determined because treatment depends on stone type (Table 279-1). No matter what disorders are

found, every patient should be counseled to avoid dehydration and to drink sufficient water so that they excrete at least 2 L of urine every day. Since treatment is prolonged, the use of medications must be justified by the activity and severity of stone disease and the importance of protection against new stones.

TREATMENT

The management of stones already present in the kidneys or urinary tract requires a combined medical and surgical approach. The specific treatment depends on the location of the stone, the extent of obstruction, the function of the affected and unaffected kidney, the presence or absence of urinary tract infection, the progress of stone passage, and the risks of operation or anesthesia given the clinical state of the patient. In general, severe obstruction, infection, intractable pain, and serious bleeding are indications for removal of a stone.

In the past, stones were removed by operation or by passing a flexible basket retrograde up the ureter from the bladder during cystoscopy. There are now three alternatives. *Extracorporeal lithotripsy* causes the in situ fragmentation of stones in the kidney, renal pelvis, or ureter by exposing them to shock waves. The kidney stone is centered at a focal point of parabolic reflectors, and high-intensity shock waves are created by high-voltage discharge. The waves are transmitted to the patient using water as a conduction medium, either by placing the patient in a water tank or by placing water-filled cushions between the patient and the shock wave generators. After multiple discharges, most stones are reduced to powder that moves through the ureter into the bladder. *Percutaneous ultrasonic lithotripsy* requires the passage of a rigid cystoscope-like instrument into the renal pelvis through a small incision in the flank. Stones can be disrupted by a small ultrasound transducer, and fragments can be removed directly. The last method is *laser lithotripsy via a ureteroscope* for removal of ureteral stones. These various forms of lithotripsy have largely replaced pyelolithotomy and ureterolithotomy.

CALCIUM STONES

Idiopathic Hypercalciuria (See also Chap. 341) This condition appears to be hereditary, and its diagnosis is straightforward (Table 279-1). In some patients, primary intestinal hyperabsorption of calcium causes transient postprandial hypercalcemia that suppresses secretion of parathyroid hormone. The renal tubules are deprived of the normal stimulus to reabsorb calcium at the same time that the filtered load of calcium is increased. In other patients, reabsorption of calcium by the renal tubules appears to be defective, and secondary hyperparathyroidism is evoked by urinary losses of calcium. Renal synthesis of 1,25-dihydroxyvitamin D is increased, enhancing intestinal absorption of calcium. In the past, the separation of "absorptive" and "renal" forms of hypercalciuria was used to guide treatment. However, these may not be distinct entities but the extremes of a continuum of behavior. Vitamin D overactivity, either through high vitamin D levels or excess vitamin D receptor, is a likely explanation for the hypercalciuria in many of these patients. Hypercalciuria contributes to stone formation by raising urine saturation with respect to calcium oxalate and calcium phosphate.

TREATMENT

Thiazide diuretics lower urine calcium in idiopathic hypercalciuria and are effective in preventing the formation of stones. Three 3-year randomized trials have shown a 50% decrease in stone formation in the thiazide-treated group as compared to the placebo-treated controls. The drug effect requires slight contraction of the extracellular fluid volume, and massive use of NaCl reduces its therapeutic effect. Thiazide-induced hypokalemia should be aggressively treated since hypokalemia will reduce urine citrate, increasing urine calcium ion levels.

Hyperuricosuria About 20% of calcium oxalate stone formers are hyperuricosuric, primarily because of an excessive intake of purine from meat, fish, and poultry. The mechanism of stone formation is probably due to salting out calcium oxalate by urate. A low-purine diet is desirable but difficult for many patients to achieve. The alternative is allopurinol, which has been shown to be effective in a randomized controlled trial. A dose of 100 mg bid is usually sufficient.

Primary Hyperparathyroidism (See also Chap. 341) The diagnosis of this condition is established by documenting that hypercalcemia that cannot be otherwise explained is accompanied by inappropriately elevated serum concentrations of parathyroid hormone. Hypercalciuria, usually present, raises the urine supersaturation of calcium phosphate and/or calcium oxalate (Table 279-1). Prompt diagnosis is important because parathyroidectomy should be carried out before renal damage or bone disease occurs.

Distal Renal Tubular Acidosis (See also Chap. 276) The defect in this condition seems to reside in the distal nephron, which cannot establish a normal pH gradient between urine and blood, leading to hyperchloremic acidosis. The diagnosis is suggested by a minimum urine pH in the presence of systemic acidosis above 5.5. If the diagnosis is in doubt because metabolic abnormalities are mild, oral challenge with NH_4Cl, 1.9 mmol/kg of body weight, will not lower urine pH below 5.5 in patients with distalRTA. Hypercalciuria, an alkaline urine, and a low urine citrate level cause supersaturation with respect to calcium phosphate. Calcium phosphate stones form, nephrocalcinosis is common, and osteomalacia or rickets may occur. Renal damage is frequent, and glomerular filtration rate falls gradually. Treatment with supplemental alkali reverses hypercalciuria and limits the production of new stones. The usual dose of sodium bicarbonate is 0.5 to 2.0 mmol/kg of body weight per day in four to six divided doses. An alternative is potassium citrate supplementation, given at the same dose per day but needing to be given only three to four times per day. In incomplete distal RTA, systemic acidosis is absent, but urine pH cannot be lowered below 5.5 after an exogenous acid load such as ammonium chloride. Incomplete RTA may develop in some patients who form calcium oxalate stones because of idiopathic hypercalciuria; the importance of RTA in producing stones in this situation is uncertain, and thiazide treatment is a reasonable alternative. Some patients with incomplete RTA form calcium phosphate stones because of low urine citrate and an alkaline urine and are best treated with alkali as if RTA were complete.

Hyperoxaluria Overabsorption of dietary oxalate and consequent oxaluria, i.e., so-called intestinal oxaluria, is one consequence of fat malabsorption (Chap. 286). The latter can be caused by a variety of conditions, including bacterial overgrowth syndromes, chronic disease of the pancreas and biliary tract, jejunoileal bypass in

treatment of obesity, or ileal resection for inflammatory bowel disease. With fat malabsorption, calcium in the bowel lumen is bound by fatty acids instead of oxalate, which is left free for absorption in the colon. Delivery of unabsorbed fatty acids and bile salts to the colon may injure the colonic mucosa and enhance oxalate absorption. Dietary excess of oxalate in patients with normal intestinal function is a common cause of mild elevation of urine oxalate, but seldom to the level seen in patients with enteric hyperoxaluria. Hereditary hyperoxaluria states are rare causes of severe hyperoxaluria; patients usually present with recurrent calcium oxalate stones during childhood. Type I hereditary hyperoxaluria is inherited as an autosomal recessive trait and is due to a deficiency in the peroxisomal enzyme alanine:glyoxylate aminotransferase. Type II is due to a deficiency of D-glyceric dehydrogenase. Ethylene glycol intoxication and methoxyflurane also can cause oxalate overproduction and hyperoxaluria. Hyperoxaluria from any cause can produce tubulointerstitial nephropathy (Chap. 277) and lead to stone formation.

TREATMENT

The oxalate-binding resin cholestyramine at a dose of 8 to 16 g/d, correction of fat malabsorption, and a low-fat, low oxalate diet are effective treatments for oxaluria secondary to intestinal overabsorption. Calcium supplements, given with meals, precipitate oxalate in the gut lumen providing an alternative form of therapy. Treatment for hereditary hyperoxaluria includes a high fluid intake, neutral phosphate, and pyridoxine (25 to 200 mg/d). Citrate supplementation may also have some benefit. Even with aggressive therapy, irreversible renal failure secondary to recurrent stone formation often occurs. Segmental liver transplant, to correct the enzyme defect, combined with a kidney transplant have been successfully utilized in patients with hereditary hyperoxaluria.

Hypocitraturia Urine citrate prevents calcium stone formation by creating a soluble complex with calcium, effectively reducing free urine calcium. Hypocitraturia is found in 15 to 60% of stone formers, either as a single disorder or in combination with other metabolic abnormalities. It can be secondary to systemic disorders, such as RTA, chronic diarrheal illness, or hypokalemia, or it may be a primary disorder, in which case it is called *idiopathic hypocitraturia*.

TREATMENT

Treatment is with alkali, which increases urine citrate excretion; generally bicarbonate or citrate salts are used. Potassium salts are preferred as sodium loading increases urinary excretion of calcium, reducing the effectiveness of treatment. A recent randomized, placebo-controlled trial has demonstrated the effectiveness of potassium citrate in idiopathic hypocitraturia.

Idiopathic Calcium Lithiasis Some patients have no metabolic cause for stones despite a thorough metabolic evaluation (Table 279-1). The best treatment appears to be high fluid intake so that the urine specific gravity remains at 1.005 or below throughout the day and night. Oral phosphate at a dose of 2 g phosphorus daily may lower urine calcium and increase urine pyrophosphate and thereby reduce the rate of recurrence. Orthophosphate causes mild nausea and diarrhea initially, but tolerance

may improve with continued intake. Thiazide treatment to reduce calcium excretion and allopurinol to diminish uric acid output also may be helpful.

URIC ACID STONES

These stones form because the urine becomes supersaturated with undissociated uric acid that is protonated at its N-9 position. In gout, idiopathic uric acid lithiasis, and dehydration, the average pH is usually below 5.4 and often below 5.0. Undissociated uric acid therefore predominates and is soluble in urine only in concentrations of 100 mg/L. Concentrations above this level represent supersaturation that causes crystals and stones to form. Hyperuricosuria, when present, increases supersaturation, but urine of low pH can be supersaturated with undissociated uric acid even though the daily excretion rate is normal. Myeloproliferative syndromes, chemotherapy of malignant tumors, and the Lesch-Nyhan syndrome cause such massive production of uric acid and consequent hyperuricosuria that stones and uric acid sludge form even at a normal urine pH. Plugging of the renal collecting tubules by uric acid crystals can cause acute renal failure.

TREATMENT

The two goals of treatment are to raise urine pH and to lower excessive urine uric acid excretion to less than 1 g/d. Supplemental alkali, 1 to 3 mmol/kg of body weight per day, should be given in three or four evenly spaced, divided doses, one of which should be given at bedtime. The form of the alkali may be important. Potassium citrate may reduce the risk of calcium salts crystallizing when urine pH is increased, whereas sodium citrate or sodium bicarbonate may increase the risk. If the overnight urine pH is below 5.5, the evening dose of alkali may be raised or 250 mg acetazolamide added at bedtime. A low-purine diet should be instituted in those uric acid stone formers with hyperuricosuria. Patients who continue to form uric acid stones despite treatment with fluids, alkali, and a low-purine diet should have allopurinol added to their regimen. If hypercalciuria is also present, it should be specifically treated, as alkali alone could lead to calcium phosphate stone formation.

CYSTINURIA AND CYSTINE STONES (See also Chap. 352)

In this disorder, proximal tubular and jejunal transport of the dibasic amino acids cystine, lysine, arginine, and ornithine are defective, and excessive amounts are lost in the urine. Clinical disease is due solely to the insolubility of cystine, which forms stones.

Pathogenesis Cystinuria occurs because of defective transport of amino acids by the brush borders of renal tubule and intestinal epithelial cells. Cystine, lysine, arginine, and ornithine appear to share a common renal transport pathway, since infusion of lysine decreases tubular reabsorption of the other three. However, cystine is also transported by a separate transport mechanism, because cystinuria and dibasic aminoaciduria can occur independently. The intestinal defects are not similar in all patients who are homozygous for cystinuria, and the extent of aminoaciduria in individuals who are heterozygous carriers of the defect varies from family to family. Three types of inheritance have been described (Chap. 352). A gene located on chromosome 2 and designated SLC3A1, codes for a dibasic amino acid transporter and has been found to

be abnormal in Type I cystinuria. Linkage analysis has mapped Type III cystinuria to chromosome 19.

Diagnosis Cystine stones are formed only by patients with cystinuria, but 10% of stones in cystinuric patients do not contain cystine; therefore, every stone former should be screened for the disease. The sediment from a first morning urine specimen in many patients with homozygous cystinuria reveals typical flat, hexagonal, platelike cystine crystals. Cystinuria also can be detected using the urine sodium nitroprusside test. The test is positive with 75 to 125 mg cystine per gram of creatinine, a concentration lower than that in the urine of patients with cystinuria but above the levels in normal urine. Because the test is sensitive, it is positive in many asymptomatic heterozygotes for cystinuria. A positive nitroprusside test or the finding of cystine crystals in the urine sediment should be evaluated by measurement of daily cystine excretion. Normal adults excrete 40 to 60 mg cystine per gram of creatinine, heterozygotes usually excrete less than 300 mg/g, and homozygotes almost always excrete above 250 mg/g.

TREATMENT

This consists of a high fluid intake, even at night. Daily urine volume should exceed 3 L. Raising urine pH with alkali is helpful, provided the urine pH exceeds 7.5. A low-salt diet (100 mmol/d) can reduce cystine excretion up to 40%. Because side effects are frequent, drugs such as penicillamine and tiopronin, which form the soluble disulfide cysteine-drug complexes, should be used only when fluid loading, salt reduction, and alkali therapy are ineffective. Captopril, which has a free sulfhydryl group to bind cysteine, has been used in a limited number of patients with some success. Low-methionine diets have not proved to be practical for clinical use, but patients should avoid protein gluttony.

STRUVITE STONES

These stones are a result of urinary infection with bacteria, usually *Proteus* species, which possess urease, an enzyme that degrades urea to NH_3 and CO_2. The NH_3hydrolyzes to $NH_4^+$and raises urine pH to 8 or 9. The CO_2hydrates to $H_2CO_3$and then dissociates to CO_3^{2-}which precipitates with calcium as $CaCO_3$. The NH_4^+precipitates $PO_4^{3-}$and $Mg^{2+}$to form $MgNH_4PO_4$(struvite). The result is a stone of calcium carbonate admixed with struvite. Struvite does not form in urine in the absence of infection, because NH_4^+concentration is low in urine that is alkaline in response to physiologic stimuli. Chronic *Proteus* infection can occur because of impaired urinary drainage, urologic instrumentation or surgery, and especially with chronic antibiotic treatment, which can favor the dominance of *Proteus* in the urinary tract.

TREATMENT

Complete removal of the stone with subsequent sterilization of the urinary tract is the treatment of choice for patients who can tolerate the procedures. Open surgery is successful in debulking the stone and improving renal function if obstruction is present; however, there is recurrence of stone in 25% of the patients. Irrigation of the renal pelvis and calyces with hemiacidrin, a solution that dissolves struvite, can reduce recurrence after surgery. Newer procedures such as lithotripsy and percutaneous nephrolithotomy,

alone or in combination, have largely replaced open surgery. Stone-free rates of 50 to 90% have been reported after these procedures. Antimicrobial treatment is best reserved for dealing with acute infection and for maintenance of a sterile urine after surgery, in the hope of preventing recurrence or minimizing stone growth. Urine cultures and culture of stone fragments removed at surgery should guide the choice of antibiotic. Methenamine mandelate, which lowers urine pH and liberates formaldehyde, can be used for chronic suppression of infection when a stone is present. For patients who are not candidates for surgical removal of stone, acetohydroxamic acid, an inhibitor of urease, can be used. Though effective in treating the stones, acetohydroxamic acid has many side effects, such as headache, tremor, and thrombophlebitis, which limits its use. Lowering urine pH with chronic administration of NH_4Cl may retard stone growth but also may raise urine calcium level and promote the formation of calcium oxalate stones.

(Bibliography omitted in Palm version)

Back to Table of Contents

280. URINARY TRACT INFECTIONS AND PYELONEPHRITIS - *Walter E. Stamm*

DEFINITIONS

Acute infections of the urinary tract can be subdivided into two general anatomic categories: lower tract infection (urethritis and cystitis) and upper tract infection (acute pyelonephritis, prostatitis, and intrarenal and perinephric abscesses). Infections at these various sites may occur together or independently and may either be asymptomatic or present as one of the clinical syndromes described below. Infections of the urethra and bladder are often considered superficial (or mucosal) infections, while prostatitis, pyelonephritis, and renal suppuration signify tissue invasion.

From a microbiologic perspective, urinary tract infection (UTI) exists when pathogenic microorganisms are detected in the urine, urethra, bladder, kidney, or prostate. In most instances, growth of more than 10^5 organisms per milliliter from a properly collected midstream "clean-catch" urine sample indicates infection. However, significant bacteriuria is lacking in some cases of true UTI. Especially in symptomatic patients, a smaller number of bacteria (10^2 to 10^4/mL) may signify infection. In urine specimens obtained by suprapubic aspiration or "in-and-out" catheterization and in samples from a patient with an indwelling catheter, colony counts of 10^2 to 10^4/mL generally indicate infection. Conversely, colony counts of $>10^5$/mL of midstream urine are occasionally due to specimen contamination, which is especially likely when multiple species are found.

Infections that recur after antibiotic therapy can be due to the persistence of the originally infecting strain (as judged by species, antibiogram, serotype, and molecular type) or to reinfection with a new strain. "Same-strain" recurrent infections that become evident within 2 weeks of cessation of therapy can be the result of unresolved renal or prostatic infection (termed *relapse*) or of persistent vaginal or intestinal colonization leading to rapid reinfection of the bladder.

Symptoms of dysuria, urgency, and frequency that are unaccompanied by significant bacteriuria have been termed the *acute urethral syndrome*. Although widely used, this term lacks anatomic precision because many cases so designated are actually bladder infections. Moreover, since the causative agent can usually be identified in these patients, the term *syndrome* -- implying unknown causation -- is inappropriate.

Chronic pyelonephritis refers to chronic interstitial nephritis believed to result from bacterial infection of the kidney (Chap. 277). Many noninfectious diseases also cause an interstitial nephritis that is indistinguishable pathologically from chronic pyelonephritis.

ACUTE UTIS: URETHRITIS, CYSTITIS, AND PYELONEPHRITIS

EPIDEMIOLOGY

Epidemiologically, UTIs are subdivided into catheter-associated (or nosocomial) infections and non-catheter-associated (or community-acquired) infections. Infections in either category may be symptomatic or asymptomatic. Acute community-acquired infections are very common and account for more than 7 million office visits annually in

the United States. These infections occur in 1 to 3% of schoolgirls and then increase markedly in incidence with the onset of sexual activity in adolescence. The vast majority of acute symptomatic infections involve young women; a prospective study demonstrated an annual incidence of 0.5 to 0.7 infections per patient-year in this group. Acute symptomatic UTIs are unusual in men under the age of 50. The development of asymptomatic bacteriuria parallels that of symptomatic infection and is rare among men under 50 but common among women between 20 and 50. Asymptomatic bacteriuria is more common among elderly men and women, with rates as high as 40 to 50% in some studies.

ETIOLOGY

Many different microorganisms can infect the urinary tract, but by far the most common agents are the gram-negative bacilli. *Escherichia coli* causes approximately 80% of acute infections in patients without catheters, urologic abnormalities, or calculi. Other gram-negative rods, especially *Proteus* and *Klebsiella* and occasionally *Enterobacter*, account for a smaller proportion of uncomplicated infections. These organisms, plus *Serratia* and *Pseudomonas*, assume increasing importance in recurrent infections and in infections associated with urologic manipulation, calculi, or obstruction. They play a major role in nosocomial, catheter-associated infections (see below). *Proteus* spp., by virtue of urease production, and *Klebsiella* spp., through the production of extracellular slime and polysaccharides, predispose to stone formation and are isolated more frequently from patients with calculi.

Gram-positive cocci play a lesser role in UTIs. However, *Staphylococcus saprophyticus* -- a novobiocin-resistant, coagulase-negative species -- accounts for 10 to 15% of acute symptomatic UTIs in young females. Enterococci occasionally cause acute uncomplicated cystitis in women. More commonly, enterococci and *Staphylococcus aureus* cause infections in patients with renal stones or previous instrumentation or surgery. Isolation of *S. aureus* from the urine should arouse suspicion of bacteremic infection of the kidney.

About one-third of women with dysuria and frequency have either an insignificant number of bacteria in midstream urine cultures or completely sterile cultures and have been previously defined as having the urethral syndrome. About three-quarters of these women have pyuria, while one-quarter have no pyuria and little objective evidence of infection. In the women with pyuria, two groups of pathogens account for most infections. Low quantities (10^2 to 10^4 bacteria per milliliter) of typical bacterial uropathogens such as *E. coli*, *S. saprophyticus*, *Klebsiella*, or *Proteus* are found in midstream urine specimens from most of these women. These bacteria are probably the causative agents in these infections because they can usually be isolated from a suprapubic aspirate, are associated with pyuria, and respond to appropriate antimicrobial therapy. In other women with acute urinary symptoms, pyuria, and urine that is sterile (even when obtained by suprapubic aspiration), sexually transmitted urethritis-producing agents such as *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, and herpes simplex virus are etiologically important. These agents are found most frequently in young, sexually active women with new sexual partners.

The causative role of nonbacterial pathogens in UTIs remains poorly defined.

Ureaplasma urealyticum has frequently been isolated from the urethra and urine of patients with acute dysuria and frequency but is also found in specimens from many patients without urinary symptoms. Ureaplasmas probably account for some cases of urethritis and cystitis. *U. urealyticum* and *Mycoplasma hominis* have been isolated from prostatic and renal tissues of patients with acute prostatitis and pyelonephritis, respectively, and are probably responsible for some of these infections as well. Adenoviruses cause acute hemorrhagic cystitis in children and in some young adults, often in epidemics. Although other viruses can be isolated from urine (e.g., cytomegalovirus), they are thought not to cause UTI. Colonization of the urine of catheterized or diabetic patients by *Candida* and other fungal species is common and sometimes progresses to symptomatic invasive infection (Chap. 205).**Mycobacterial infection of the genitourinary tract is discussed in Chap. 169.*

PATHOGENESIS AND SOURCES OF INFECTION

The urinary tract should be viewed as a single anatomic unit that is united by a continuous column of urine extending from the urethra to the kidney. In the vast majority ofUTIs, bacteria gain access to the bladder via the urethra. Ascent of bacteria from the bladder may follow and is probably the pathway for most renal parenchymal infections.

The vaginal introitus and distal urethra are normally colonized by diphtheroids, streptococcal species, lactobacilli, and staphylococcal species but not by the enteric gram-negative bacilli that commonly causeUTIs. In females prone to the development of cystitis, however, enteric gram-negative organisms residing in the bowel colonize the introitus, the periurethral skin, and the distal urethra before and during episodes of bacteriuria. The factors that predispose to periurethral colonization with gram-negative bacilli remain poorly understood, but alteration of the normal vaginal flora by antibiotics, other genital infections, or contraceptives (especially spermicide) appears to play an important role. Loss of the normally dominant H_2O_2-producing lactobacilli in the vaginal flora appears to facilitate colonization by *E. coli.* Small numbers of periurethral bacteria probably gain entry to the bladder frequently, a process that is facilitated in some cases by urethral massage during intercourse. Whether bladder infection ensues depends on interacting effects of the pathogenicity of the strain, the inoculum size, and the local and systemic host defense mechanisms.

Under normal circumstances, bacteria placed in the bladder are rapidly cleared, partly through the flushing and dilutional effects of voiding but also as a result of the antibacterial properties of urine and the bladder mucosa. Owing mostly to a high urea concentration and high osmolarity, the bladder urine of many normal persons inhibits or kills bacteria. Prostatic secretions possess antibacterial properties as well. Polymorphonuclear leukocytes enter the bladder epithelium and the urine soon after infection arises and play a role in clearing bacteriuria. The role of locally produced antibody remains unclear.

Hematogenous pyelonephritis occurs most often in debilitated patients who are either chronically ill or receiving immunosuppressive therapy. Metastatic staphylococcal or candidal infections of the kidney may follow bacteremia or fungemia, spreading from distant foci of infection in the bone, skin, vasculature, or elsewhere.

CONDITIONS AFFECTING PATHOGENESIS

Gender and Sexual Activity The female urethra appears to be particularly prone to colonization with colonic gram-negative bacilli because of its proximity to the anus, its short length (about 4 cm), and its termination beneath the labia. Sexual intercourse causes the introduction of bacteria into the bladder and is temporally associated with the onset of cystitis; it thus appears to be important in the pathogenesis ofUTIs in younger women. Voiding after intercourse reduces the risk of cystitis, probably because it promotes the clearance of bacteria introduced during intercourse. In addition, use of spermicidal compounds with a diaphragm or cervical cap or of spermicide-coated condoms dramatically alters the normal introital bacterial flora and has been associated with marked increases in vaginal colonization with *E. coli* and in the risk of UTI.

In males who are <50 years old and who have no history of heterosexual or homosexual rectal intercourse,UTI is exceedingly uncommon, and this diagnosis should be questioned in the absence of clear documentation. An important factor predisposing to bacteriuria in men is urethral obstruction due to prostatic hypertrophy. Homosexuality is also associated with an increased risk of cystitis in men, probably related to rectal intercourse. Men (and women) who are infected with HIV and who have CD4+ T cell counts of <200/uL are at increased risk of both bacteriuria and symptomatic UTI. Finally, lack of circumcision has been identified as a risk factor for UTI in both neonates and young men.

Pregnancy UTIs are detected in 2 to 8% of pregnant women. Symptomatic upper tract infections, in particular, are unusually common during pregnancy; fully 20 to 30% of pregnant women with asymptomatic bacteriuria subsequently develop pyelonephritis. This predisposition to upper tract infection during pregnancy results from decreased ureteral tone, decreased ureteral peristalsis, and temporary incompetence of the vesicoureteral valves. Bladder catheterization during or after delivery causes additional infections. Increased incidences of low-birth-weight infants, premature delivery, and newborn mortality result from UTIs during pregnancy, particularly those infections involving the upper tract.

Obstruction Any impediment to the free flow of urine -- tumor, stricture, stone, or prostatic hypertrophy -- results in hydronephrosis and a greatly increased frequency ofUTI. Infection superimposed on urinary tract obstruction may lead to rapid destruction of renal tissue. It is of utmost importance, therefore, when infection is present, to identify and repair obstructive lesions. On the other hand, when an obstruction is minor and is not progressive or associated with infection, great caution should be exercised in attempting surgical correction. The introduction of infection in such cases may be more damaging than an uncorrected minor obstruction that does not significantly impair renal function.

Neurogenic Bladder Dysfunction Interference with the nerve supply to the bladder, as in spinal cord injury, tabes dorsalis, multiple sclerosis, diabetes, and other diseases, may be associated withUTI. The infection may be initiated by the use of catheters for bladder drainage and is favored by the prolonged stasis of urine in the bladder. An additional factor often operative in these cases is bone demineralization due to immobilization, which causes hypercalciuria, calculus formation, and obstructive

uropathy.

Vesicoureteral Reflux Defined as reflux of urine from the bladder cavity up into the ureters and sometimes into the renal pelvis, vesicoureteral reflux occurs during voiding or with elevation of pressure in the bladder. In practice, this condition is demonstrated by the finding of retrograde movement of radiopaque or radioactive material during a voiding cystourethrogram. An anatomically impaired vesicoureteral junction facilitates reflux of bacteria and thus upper tract infection. However, since a fluid connection between the bladder and the kidney always exists, even in the normal urinary system, some retrograde movement of bacteria probably takes place during infection but is not detected by radiologic techniques.

Vesicoureteral reflux is common among children with anatomic abnormalities of the urinary tract as well as among children with anatomically normal but infected urinary tracts. In the latter group, reflux disappears with advancing age and is probably attributable to factors other thanUTI. Long-term follow-up of children with UTI who have reflux has established that renal damage correlates with marked reflux, not with infection.

The routine search for reflux would be aided by the development of noninvasive tests applicable to young children, in whom the need for an effective technique is greatest. In the meantime, it appears reasonable to search for reflux in anyone with unexplained failure of renal growth or with renal scarring, becauseUTI per se is an insufficient explanation for these abnormalities. On the other hand, it is doubtful that all children who have recurrent UTIs but whose urinary tract appears normal on pyelography should be subjected to voiding cystoureterography merely for the detection of the rare patient with marked reflux not revealed by the intravenous pyelogram.

Bacterial Virulence Factors Not all strains of *E. coli* are equally capable of infecting the intact urinary tract. Bacterial virulence factors markedly influence the likelihood that a given strain, once introduced into the bladder, will causeUTI. Most *E. coli* strains that cause symptomatic UTIs in noncatheterized patients belong to a small number of specific O, K, and H serogroups. These uropathogenic clones have accumulated a number of virulence genes that are often closely linked on the bacterial chromosome in "virulence islands." Adherence of bacteria to uroepithelial cells is a critical first step in the initiation of infection. For both *E. coli* and *Proteus*, fimbriae (hairlike proteinaceous surface appendages) mediate the attachment of bacteria to specific receptors on epithelial cells. The attachment of bacteria to uroepithelial cells initiates a number of important events in the mucosal epithelial cell, including secretion of interleukin (IL) 6 and IL-8 and induction of both apoptosis and epithelial cell desquamation. Besides fimbriae, uropathogenic *E. coli* strains usually produce hemolysin and aerobactin (a siderophore for scavenging iron) and are resistant to the bactericidal action of human serum. Nearly all *E. coli* strains causing acute pyelonephritis and most of those causing acute cystitis are uropathogenic. In contrast, infections in patients with structural or functional abnormalities of the urinary tract are generally caused by bacterial strains that lack these uropathogenic properties; the implication is that these properties are not needed for infection of the compromised urinary tract.

Genetic Factors Increasing evidence suggests that host genetic factors influence

susceptibility toUTI. A maternal history of UTI is more often found among women who have experienced recurrent UTIs than among controls. The number and type of receptors on uroepithelial cells to which bacteria may attach are at least in part genetically determined. Many of these structures are components of blood group antigens and are present on both erythrocytes and uroepithelial cells. For example, P fimbriae mediate attachment of *E. coli* to P-positive erythrocytes and are found on nearly all strains causing acute uncomplicated pyelonephritis. Conversely, P blood group-negative individuals, who lack these receptors, have a decreased likelihood of pyelonephritis. It has also been demonstrated that nonsecretors of blood group antigens are at increased risk of recurrent UTI; this predisposition may relate to a different profile of genetically determined glycolipids on uroepithelial cells.

LOCALIZATION OF INFECTION

Unfortunately, currently available methods of distinguishing renal parenchymal infection from cystitis are neither reliable nor convenient enough for routine clinical use. Fever or an elevated level of C-reactive protein often accompanies acute pyelonephritis and is found in rare cases of cystitis but may also occur in infections other than pyelonephritis.

CLINICAL PRESENTATION

Cystitis Patients with cystitis usually report dysuria, frequency, urgency, and suprapubic pain. The urine often becomes grossly cloudy and malodorous, and it is bloody in about 30% of cases. White cells and bacteria can be detected by examination of unspun urine in most cases. However, some women with cystitis have only 10^2 to 10^4bacteria per milliliter of urine, and in these instances bacteria cannot be seen in a Gram-stained preparation of unspun urine. Physical examination generally reveals only tenderness of the urethra or the suprapubic area. If a genital lesion or a vaginal discharge is evident, especially in conjunction with fewer than 10^5bacteria per milliliter on urine culture, then pathogens that may cause urethritis, vaginitis, or cervicitis, such as *C. trachomatis*, *N. gonorrhoeae*, *Trichomonas*, *Candida*, and herpes simplex virus, should be considered. Prominent systemic manifestations such as a temperature of>38.3°C (>101°F), nausea, and vomiting usually indicate concomitant renal infection, as does costovertebral angle tenderness. However, the absence of these findings does not ensure that infection is limited to the bladder and urethra.

Acute Pyelonephritis Symptoms of acute pyelonephritis generally develop rapidly over a few hours or a day and include a fever, shaking chills, nausea, vomiting, and diarrhea. Symptoms of cystitis may or may not be present. Besides fever, tachycardia, and generalized muscle tenderness, physical examination reveals marked tenderness on deep pressure in one or both costovertebral angles or on deep abdominal palpation. In some patients, signs and symptoms of gram-negative sepsis predominate. Most patients have significant leukocytosis and bacteria detectable in Gram-stained unspun urine. Leukocyte casts are present in the urine of some patients, and the detection of these casts is pathognomonic. Hematuria may be demonstrated during the acute phase of the disease; if it persists after acute manifestations of infection have subsided, a stone, a tumor, or tuberculosis should be considered.

Except in individuals with papillary necrosis, abscess formation, or urinary obstruction,

the manifestations of acute pyelonephritis usually respond to therapy within 48 to 72 h. However, despite the absence of symptoms, bacteriuria or pyuria may persist. In severe pyelonephritis, fever subsides more slowly and may not disappear for several days, even after appropriate antibiotic treatment has been instituted.

Urethritis Approximately 30% of women with acute dysuria, frequency, and pyuria have midstream urine cultures that show either no growth or insignificant bacterial growth. Clinically, these women cannot always be readily distinguished from those with cystitis. In this situation, a distinction should be made between women infected with sexually transmitted pathogens, such as *C. trachomatis*, *N. gonorrhoeae*, or herpes simplex virus, and those with low-count *E. coli* or staphylococcal infection of the urethra and bladder. Chlamydial or gonococcal infection should be suspected in women with a gradual onset of illness, no hematuria, no suprapubic pain, and >7 days of symptoms. The additional history of a recent sex-partner change, especially if the patient's partner has recently had chlamydial or gonococcal urethritis, should heighten the suspicion of a sexually transmitted infection, as should the finding of mucopurulent cervicitis (Chaps. 132 and133). Gross hematuria, suprapubic pain, an abrupt onset of illness, a duration of illness of <3 days, and a history ofUTIsfavor the diagnosis of *E. coli* UTI.

Catheter-Associated UTIs (See alsoChap. 135) Bacteriuria develops in at least 10 to 15% of hospitalized patients with indwelling urethral catheters. The risk of infection is about 3 to 5% per day of catheterization. *E. coli*, *Proteus*, *Pseudomonas*, *Klebsiella*, *Serratia*, staphylococci, enterococci, and *Candida* usually cause these infections. Many infecting strains display markedly greater antimicrobial resistance than organisms that cause community-acquiredUTIs. Factors associated with an increased risk of catheter-associated UTI include female sex, prolonged catheterization, severe underlying illness, disconnection of the catheter and drainage tube, other types of faulty catheter care, and lack of systemic antimicrobial therapy.

Infection occurs when bacteria reach the bladder by one of two routes: by migrating through the column of urine in the catheter lumen (intraluminal route) or by moving up the mucous sheath outside the catheter (periurethral route). Hospital-acquired pathogens reach the patient's catheter or urine-collecting system on the hands of hospital personnel, in contaminated solutions or irrigants, and via contaminated instruments or disinfectants. Bacteria usually enter the catheter system at the catheter-collecting tube junction or at the drainage bag portal. The organisms then ascend intraluminally into the bladder within 24 to 72 h. Alternatively, the patient's own bowel flora may colonize the perineal skin and periurethral area and reach the bladder via the external surface of the catheter. This route is particularly common in women. Studies have demonstrated the importance of the attachment and growth of bacteria on the surfaces of the catheter in the pathogenesis of catheter-associatedUTI. Such bacteria growing in biofilms on the catheter eventually produce encrustations consisting of bacteria, bacterial glycocalyces, host urinary proteins, and urinary salts. These encrustations provide a refuge for bacteria and may protect them from antimicrobial agents and phagocytes.

Clinically, most catheter-associated infections cause minimal symptoms and no fever and often resolve after withdrawal of the catheter. The frequency of upper tract infection associated with catheter-induced bacteriuria is unknown. Gram-negative bacteremia,

which follows catheter-associated bacteriuria in 1 to 2% of cases, is the most significant recognized complication of catheter-inducedUTIs. The catheterized urinary tract has repeatedly been demonstrated to be the most common source of gram-negative bacteremia in hospitalized patients, generally accounting for about 30% of cases.

Catheter-associatedUTIs can sometimes be prevented in patients catheterized for<2 weeks by use of a sterile closed collecting system, by attention to aseptic technique during insertion and care of the catheter, and by measures to minimize cross-infection. Other preventive approaches, including short courses of systemic antimicrobial therapy, topical application of periurethral antimicrobial ointments, use of preconnected catheter-drainage tube units, use of catheters impregnated with antimicrobial agents, and addition of antimicrobial drugs to the drainage bag, have all been protective in at least one controlled trial but are not recommended for general use. Despite precautions, the majority of patients catheterized for>2 weeks eventually develop bacteriuria. The need for treatment as well as the optimal type and duration of treatment for such patients with asymptomatic bacteriuria have not been established. Removal of the catheter in conjunction with a short course of antibiotics to which the organism is susceptible probably constitutes the best course of action and nearly always eradicates bacteriuria. Treatment of asymptomatic catheter-associated bacteriuria may be of greatest benefit to elderly women, who most often develop symptoms if left untreated. If the catheter cannot be removed, antibiotic therapy usually proves to be unsuccessful and may in fact result in infection with a more resistant strain. In this situation, the bacteriuria should be ignored unless the patient develops symptoms or is at high risk of developing bacteremia. In these cases, use of systemic antibiotics or urinary bladder antiseptics may reduce the degree of bacteriuria and the likelihood of bacteremia. Because of spinal cord injury, incontinence, or other factors, some patients in hospitals or nursing homes require long-term or semipermanent bladder catheterization. Measures intended to prevent infection have been largely unsuccessful, and essentially all such chronically catheterized patients develop bacteriuria. If feasible, intermittent catheterization by a nurse or by the patient appears to reduce the incidence of bacteriuria and associated complications in such patients. Treatment should be provided when symptomatic infections arise, but treatment of asymptomatic bacteriuria in such patients has no apparent benefit.

DIAGNOSTIC TESTING

Determination of the number and type of bacteria in the urine is an extremely important diagnostic procedure. In symptomatic patients, bacteria are usually present in the urine in large numbers ($\geq 10^5$/mL). In asymptomatic patients, two consecutive urine specimens should be examined bacteriologically before therapy is instituted, and$\geq 10^5$bacteria of a single species per milliliter should be demonstrable in both specimens. Since the large number of bacteria in the bladder urine is due in part to bacterial multiplication during residence in the bladder cavity, samples of urine from the ureters or renal pelvis may contain$<10^5$bacteria per milliliter and yet indicate infection. Similarly, the presence of bacteriuria of any degree in suprapubic aspirates or of$\geq 10^2$bacteria per milliliter of urine obtained by catheterization usually indicates infection. In some circumstances (antibiotic treatment, high urea concentration, high osmolarity, low pH), urine inhibits bacterial multiplication, resulting in relatively low bacterial colony counts despite infection. For this reason, antiseptic solutions should not be used in washing the periurethral area before

collection of the urine specimen. Water diuresis or recent voiding also reduces bacterial counts in urine.

Rapid methods of detection of bacteriuria have been developed as alternatives to standard culture methods. These methods detect bacterial growth by photometry, bioluminescence, or other means and provide results rapidly, usually in 1 to 2 h. Compared with urine cultures, these techniques generally exhibit a sensitivity of 95 to 98% and a negative predictive value of >99% when bacteriuria is defined as 10^5colony-forming units per milliliter. However, the sensitivity of these tests falls to 60 to 80% when 10^2 to 10^4colony-forming units per milliliter is the standard of comparison.

Microscopy of urine from symptomatic patients can be of great diagnostic value. Microscopic bacteriuria, which is best assessed with Gram-stained uncentrifuged urine, is found in more than 90% of specimens from patients whose infections are associated with colony counts of at least 10^5/mL, and this finding is very specific. However, bacteria cannot usually be detected microscopically in infections with lower colony counts (10^2 to 10^4/mL). The detection of bacteria by urinary microscopy thus constitutes firm evidence of infection, but the absence of microscopically detectable bacteria does not exclude the diagnosis. When carefully sought by means of chamber-count microscopy, pyuria is a highly sensitive indicator ofUTI in symptomatic patients. Pyuria is demonstrated in nearly all acute bacterial UTIs, and its absence calls the diagnosis into question. The leukocyte esterase "dipstick" method is less sensitive than microscopy in identifying pyuria but is a useful alternative where microscopy is not feasible. Pyuria in the absence of bacteriuria (sterile pyuria) may indicate infection with unusual bacterial agents such as *C. trachomatis*, *U. urealyticum*, and *Mycobacterium tuberculosis* or with fungi. Alternatively, sterile pyuria may be demonstrated in noninfectious urologic conditions such as calculi, anatomic abnormality, nephrocalcinosis, vesicoureteral reflux, interstitial nephritis, or polycystic disease.

Although many authorities have recommended that urine culture and antimicrobial susceptibility testing be performed for any patient with a suspectedUTI, it may be more practical and cost-effective to manage women who have symptoms characteristic of acute uncomplicated cystitis without an initial urine culture. Two approaches to presumptive therapy have generally been used. In the first, treatment is initiated solely on the basis of a typical history and/or typical findings on physical examination. In the second, women with symptoms and signs of acute cystitis and without complicating factors are managed with urinary microscopy (or, alternatively, with a leukocyte esterase test). A positive result for pyuria and/or bacteriuria provides enough evidence of infection to indicate that urine culture and susceptibility testing can be omitted and the patient treated empirically. Urine should be cultured, however, when a woman's symptoms and urine-examination findings leave the diagnosis of cystitis in question. Pretherapy cultures and susceptibility testing are also essential in the management of all patients with suspected upper tract infections and of those with complicating factors, as in these situations any of a variety of pathogens may be involved and antibiotic therapy is best tailored to the individual organism.

TREATMENT

The following principles underlie the treatment ofUTIs:

1. Except in acute uncomplicated cystitis in women, a quantitative urine culture, a Gram stain, or an alternative rapid diagnostic test should be performed to confirm infection before treatment is begun. When culture results become available, antimicrobial sensitivity testing should be used to direct therapy.

2. Factors predisposing to infection, such as obstruction and calculi, should be identified and corrected if possible.

3. Relief of clinical symptoms does not always indicate bacteriologic cure.

4. Each course of treatment should be classified after its completion as a failure (symptoms and/or bacteriuria not eradicated during therapy or in the immediate posttreatment culture) or a cure (resolution of symptoms and elimination of bacteriuria). Recurrent infections should be classified as same-strain or different-strain and as early (occurring within 2 weeks of the end of therapy) or late.

5. In general, uncomplicated infections confined to the lower urinary tract respond to short courses of therapy, while upper tract infections require longer treatment. After therapy, early recurrences due to the same strain may result from an unresolved upper tract focus of infection but often (especially after short-course therapy for cystitis) result from persistent vaginal colonization. Recurrences>2 weeks after the cessation of therapy nearly always represent reinfection with a new strain or with the previously infecting strain that has persisted in the vaginal and rectal flora.

6. Despite increasing resistance, community-acquired infections, especially initial infections, are usually due to more antibiotic-sensitive strains.

7. In patients with repeated infections, instrumentation, or recent hospitalization, the presence of antibiotic-resistant strains should be suspected.

The anatomic location of a UTI greatly influences the success or failure of a therapeutic regimen. Bladder bacteriuria (cystitis) can usually be eliminated with nearly any antimicrobial agent to which the infecting strain is sensitive; in the past, it was demonstrated that as little as a single dose of 500 mg of intramuscular kanamycin eliminated bladder bacteriuria in most cases. With upper tract infections, however, single-dose therapy fails in the majority of cases, and even a 7-day course is unsuccessful in many instances. Longer periods of treatment (2 to 6 weeks) aimed at eradicating a persistent focus of infection may be necessary in some cases.

In *acute uncomplicated cystitis*, more than 90 to 95% of infections are due to one of two organisms: *E. coli* or *S. saprophyticus*. Although resistance patterns vary geographically and resistance has increased in many areas, most strains are sensitive to many antibiotics. In most parts of the United States, more than one-quarter of *E. coli* strains causing acute cystitis are resistant to amoxicillin, sulfa drugs, and cephalexin, and resistance to trimethoprim (TMP) and trimethoprim-sulfamethoxazole (TMP-SMZ) is now approaching these levels as well.

Many have advocated single-dose treatment for acute cystitis. The advantages of

single-dose therapy include less expense, ensured compliance, fewer side effects, and perhaps less intense pressure favoring the selection of resistant organisms in the intestinal, vaginal, or perineal flora. However, more frequent recurrences develop shortly after single-dose therapy than after 3-day treatment, and single-dose therapy does not eradicate vaginal colonization with *E. coli* as effectively as do longer regimens. A 3-day course of therapy withTMP-SMZ,TMP, norfloxacin, ciprofloxacin, or ofloxacin appears to preserve the low rate of side effects of single-dose therapy while improving efficacy (Table 280-1); thus 3-day regimens are currently preferred for acute cystitis. Neither single-dose nor 3-day therapy should be used for women with symptoms or signs of pyelonephritis, urologic abnormalities or stones, or previous infections due to antibiotic-resistant organisms. Males with UTIoften have urologic abnormalities or prostatic involvement and hence are not candidates for single-dose or 3-day therapy. For empirical therapy, they should generally receive a 7- to 14-day course of a fluoroquinolone (Table 280-1).

The choice of treatment for women with acute urethritis depends on the etiologic agent involved. In chlamydial infection, azithromycin (1 g in a single oral dose) or doxycycline (100 mg orally bid for 7 days) should be used. Women with acute dysuria and frequency, negative urine cultures, and no pyuria usually do not respond to antimicrobial agents.

In women, *acute uncomplicated pyelonephritis* without accompanying clinical evidence of calculi or urologic disease is due to *E. coli* in most cases. Although the optimal route and duration of therapy have not been established, a 7- to 14-day course of a fluoroquinolone, an aminoglycoside, or a third-generation cephalosporin is usually adequate. Neither ampicillin norTMP-SMZshould be used as initial therapy because >25% of strains of *E. coli* causing pyelonephritis are now resistant to these drugs in vitro. For at least the first few days of treatment, antibiotics should probably be given intravenously to most patients, but patients with mild symptoms can be treated for 7 to 14 days with an oral antibiotic (usually ciprofloxacin or ofloxacin), with or without an initial single parenteral dose (Table 280-1). Patients who fail to respond to treatment within 72 h or who relapse after therapy should be evaluated for unrecognized suppurative foci, calculi, or urologic disease.

ComplicatedUTIs(those arising in a setting of catheterization, instrumentation, urologic anatomic or functional abnormalities, stones, obstruction, immunosuppression, renal disease, or diabetes) are typically due to hospital-acquired bacteria, including *E. coli*, *Klebsiella*, *Proteus*, *Serratia*, *Pseudomonas*, enterococci, and staphylococci. Many of the infecting strains are antibiotic-resistant. Empirical antibiotic therapy ideally provides broad-spectrum coverage against these pathogens. In patients with minimal or mild symptoms, oral therapy with a fluoroquinolone, such as ciprofloxacin or ofloxacin, can be administered until culture results and antibiotic sensitivities are known. In patients with more severe illness, including acute pyelonephritis or suspected urosepsis, hospitalization and parenteral therapy should be undertaken. Commonly used empirical regimens include imipenem alone, a penicillin or cephalosporin plus an aminoglycoside, and (when the involvement of enterococci is unlikely) ceftriaxone or ceftazidime. When information on the antimicrobial sensitivity pattern of the infecting strain becomes available, a more specific antimicrobial regimen can be selected. Therapy should generally be administered for 10 to 21 days, with the exact duration depending on the

severity of the infection and the susceptibility of the infecting strain. Follow-up cultures 2 to 4 weeks after cessation of therapy should be performed to demonstrate cure.

In *pregnancy*, acute cystitis can be managed with 7 days of treatment with amoxicillin, nitrofurantoin, or a cephalosporin. All pregnant women should be screened for asymptomatic bacteriuria during the first trimester and, if bacteriuric, should be treated with one of the regimens listed inTable 280-1. After treatment, a culture should be performed to ensure cure, and cultures should be repeated monthly thereafter until delivery. Acute pyelonephritis in pregnancy should be managed with hospitalization and parenteral antibiotic therapy, generally with a cephalosporin or an extended-spectrum penicillin. Continuous low-dose prophylaxis with nitrofurantoin should be given to women who have recurrent infections during pregnancy.

Asymptomatic bacteriuria is common, especially among elderly patients, but has not been linked to adverse outcomes in most circumstances other than pregnancy (see above). Thus antimicrobial therapy is unnecessary and may in fact promote the emergence of resistant strains in most patients with asymptomatic bacteriuria. High-risk patients with neutropenia, renal transplants, obstruction, or other complicating conditions may require treatment when asymptomatic bacteriuria occurs. Seven days of therapy with an oral agent to which the organism is sensitive should be given initially. If bacteriuria persists, it can be monitored without further treatment in most patients. Longer-term therapy (4 to 6 weeks) may be necessary in high-risk patients with persistent asymptomatic bacteriuria.

UROLOGIC EVALUATION

Very few women with recurrentUTIshave correctable lesions discovered at cystoscopy or upon intravenous pyelography, and these procedures should not be undertaken routinely in such cases. Urologic evaluation should be performed in selected instances -- namely, in women with relapsing infection, a history of childhood infections, stones or painless hematuria, or recurrent pyelonephritis. Most males with UTI should be considered to have complicated infection and thus should be evaluated urologically. Possible exceptions include young men who have cystitis associated with sexual activity, who are uncircumcised, or who have AIDS. Men or women presenting with acute infection and signs or symptoms suggestive of an obstruction or stones should undergo prompt urologic evaluation, generally by means of ultrasound.

PROGNOSIS

In patients with uncomplicated cystitis or pyelonephritis, treatment ordinarily results in complete resolution of symptoms. Lower tract infections in women are of concern mainly because they cause discomfort, morbidity, loss of time from work, and substantial health-care costs. Cystitis may also result in upper tract infection or in bacteremia (especially during instrumentation), but little evidence suggests that renal impairment follows. When repeated episodes of cystitis occur, they are nearly always reinfections, not relapses.

Acute uncomplicated pyelonephritis in adults rarely progresses to renal functional impairment and chronic renal disease. Repeated upper tract infections often represent

relapse rather than reinfection, and a vigorous search for renal calculi or an underlying urologic abnormality should be undertaken. If neither is found, 6 weeks of chemotherapy may be useful in eradicating an unresolved focus of infection.

Repeated symptomaticUTIs in children and in adults with obstructive uropathy, neurogenic bladder, structural renal disease, or diabetes progress to chronic renal disease with unusual frequency. Asymptomatic bacteriuria in these groups as well as in adults without urologic disease or obstruction predisposes to increased numbers of episodes of symptomatic infection but does not result in renal impairment in most instances.

PREVENTION

Women who experience frequent symptomaticUTIs(³3 per year on average) are candidates for long-term administration of low-dose antibiotics directed at preventing recurrences. Such women should be advised to avoid spermicide use and to void soon after intercourse. Daily or thrice-weekly administration of a single dose ofTMP-SMZ(80/400 mg),TMPalone (100 mg), or nitrofurantoin (50 mg) has been particularly effective. Norfloxacin and other fluoroquinolones have also been used for prophylaxis. Prophylaxis should be initiated only after bacteriuria has been eradicated with a full-dose treatment regimen. The same prophylactic regimens can be used after sexual intercourse to prevent episodes of symptomatic infection in women in whom UTIs are temporally related to intercourse. Other patients for whom prophylaxis appears to have some merit include men with chronic prostatitis; patients undergoing prostatectomy, both during the operation and in the postoperative period; and pregnant women with asymptomatic bacteriuria. All pregnant women should be screened for bacteriuria in the first trimester and should be treated if bacteriuria is demonstrated.

PAPILLARY NECROSIS

When infection of the renal pyramids develops in association with vascular diseases of the kidney or with urinary tract obstruction, renal papillary necrosis is likely to result. Patients with diabetes, sickle cell disease, chronic alcoholism, and vascular disease seem peculiarly susceptible to this complication. Hematuria, pain in the flank or abdomen, and chills and fever are the most common presenting symptoms. Acute renal failure with oliguria or anuria sometimes develops. Rarely, sloughing of a pyramid may take place without symptoms in a patient with chronicUTI, and the diagnosis is made when the necrotic tissue is passed in the urine or identified as a "ring shadow" on pyelography. If renal function deteriorates suddenly in a diabetic individual or a patient with chronic obstruction, the diagnosis of renal papillary necrosis should be entertained, even in the absence of fever or pain. Renal papillary necrosis is often bilateral; when it is unilateral, however, nephrectomy may be a life-saving approach to the management of overwhelming infection.

EMPHYSEMATOUS PYELONEPHRITIS AND CYSTITIS

These unusual clinical entities almost always occur in diabetic patients, often in concert with urinary obstruction and chronic infection. Emphysematous pyelonephritis is usually characterized by a rapidly progressive clinical course, with high fever, leukocytosis,

renal parenchymal necrosis, and accumulation of fermentative gases in the kidney and perinephric tissues. Most patients also have pyuria and glucosuria. *E. coli* causes most cases, but occasionally other Enterobacteriaceae are isolated. Gas in tissues can often be seen on plain films and can best be confirmed and localized by computed tomography. Surgical resection of the involved tissue in addition to systemic antimicrobial therapy is usually needed to prevent mortality in emphysematous pyelonephritis.

Emphysematous cystitis also occurs primarily in diabetic patients, usually in association with *E. coli* or other facultative gram-negative rods and often in relation to bladder outlet obstruction. Patients with this condition are generally less severely ill and have less rapidly progressive disease than those with emphysematous pyelonephritis. The patient typically reports abdominal pain, dysuria, frequency, and (in some cases) pneumaturia. Computed tomography shows gas within both the bladder lumen and the bladder wall. Generally, conservative therapy with systemic antimicrobial agents and relief of outlet obstruction are effective, but some patients do not respond to these measures and require cystectomy.

RENAL AND PERINEPHRIC ABSCESS

SeeChap. 130.

PROSTATITIS

The term *prostatitis* has been used for various inflammatory conditions affecting the prostate, including acute and chronic infections with specific bacteria and, more commonly, instances in which signs and symptoms of prostatic inflammation are present but no specific organisms can be detected. Patients with acute bacterial prostatitis can usually be identified on the basis of typical symptoms and signs, pyuria, and bacteriuria. To classify a patient with suspected chronic prostatitis correctly, first-void and midstream urine specimens, a prostatic expressate, and a postmassage urine specimen should be quantitatively cultured and evaluated for numbers of leukocytes. On the basis of the results of these studies, patients can be classified as having chronic bacterial prostatitis, chronic nonbacterial prostatitis, or prostatodynia. Patients with suspected chronic prostatitis usually have low back pain, perineal or testicular discomfort, mild dysuria, and lower urinary obstructive symptoms. Microscopic pyuria may be the only objective manifestation of prostatic disease.

ACUTE BACTERIAL PROSTATITIS

When it occurs spontaneously, this disease generally affects young men; however, it may also be associated with an indwelling urethral catheter. It is characterized by fever, chills, dysuria, and a tense or boggy, extremely tender prostate. Although prostatic massage usually produces purulent secretions with a large number of bacteria on culture, bacteremia may result from manipulation of the inflamed gland. For this reason and because the etiologic agent can usually be identified by Gram's staining and culture of urine, vigorous prostatic massage should be avoided. In non-catheter-associated cases, the infection is generally due to common gram-negative urinary tract pathogens (*E. coli* or *Klebsiella*). Initially, an intravenous fluoroquinolone, third-generation

cephalosporin, or aminoglycoside can be administered if gram-negative rods are visible in urine, and a cephalosporin or nafcillin can be given if gram-positive cocci are detected. Although many of these drugs do not readily diffuse into the noninflamed prostate gland, the response to antibiotics in acute bacterial prostatitis is usually prompt, perhaps because drugs penetrate more readily into the acutely inflamed prostate. In catheter-associated cases, the spectrum of etiologic agents is broader, including hospital-acquired gram-negative rods and enterococci. The urinary Gram stain may be particularly helpful in such cases. Imipenem, an aminoglycoside, a fluoroquinolone, or a third-generation cephalosporin should be used for initial therapy until the organism has been isolated and its susceptibilities have been determined. The long-term prognosis is good, although in some instances acute infection may result in abscess formation, epididymoorchitis, seminal vesiculitis, septicemia, and residual chronic bacterial prostatitis. Since the advent of antibiotics, the frequency of acute bacterial prostatitis has diminished markedly. In many instances, infections diagnosed as acute prostatitis are probably cases of posterior urethritis.

CHRONIC BACTERIAL PROSTATITIS

This entity is now infrequent but should be considered in men with a history of recurrent bacteriuria. Symptoms are often lacking between episodes, and the prostate usually feels normal on palpation. Obstructive symptoms or perineal pain develops in some patients. Intermittently, infection spreads to the bladder, producing frequency, urgency, and dysuria. A pattern of relapsing infection in a middle-aged man strongly suggests chronic bacterial prostatitis. Classically, the diagnosis is established by culture of *E. coli*, *Klebsiella*, *Proteus*, or other uropathogenic bacteria from the expressed prostatic secretion or postmassage urine in higher quantities than are found in first-void or midstream urine. Antibiotics promptly relieve the symptoms associated with acute exacerbations but have been less effective in eradicating the focus of chronic infection in the prostate. The relative ineffectiveness of antimicrobial agents for long-term cure results in part from the poor penetration of the prostate by most of these drugs; the low pH that prevails in this organ precludes the passage of most agents. Fluoroquinolones, including ciprofloxacin and ofloxacin, have been considerably more successful than other antimicrobials, but they must generally be given for at least 12 weeks to be effective. Patients with frequent episodes of acute cystitis in whom attempts at curative therapy fail can be managed with prolonged courses of antimicrobials (usually a sulfonamide,TMP, or nitrofurantoin), with a view toward suppressing symptoms and keeping the bladder urine sterile. Total prostatectomy obviously results in the cure of chronic prostatitis but is associated with considerable morbidity. Transurethral prostatectomy is safer but cures only one-third of patients.

NONBACTERIAL PROSTATITIS

Patients who present with symptoms and signs of prostatitis, increased numbers of leukocytes in expressed prostatic secretion and postmassage urine, no bacterial growth in cultures, and no history of recurrent episodes of bacterial prostatitis are classified as having nonbacterial prostatitis. Prostatic inflammation can be considered present when the expressed prostatic secretion and postmassage urine contain at least tenfold more leukocytes than the first-void and midstream urine specimens or when the expressed prostatic secretion contains^{3}1000 leukocytes per microliter.

The presumably infectious etiology of this condition remains unidentified. Evidence for a causative role of both *U. urealyticum* and *C. trachomatis* has been presented but is not conclusive. Since most cases of nonbacterial prostatitis occur in young, sexually active men and since many cases follow an episode of nonspecific urethritis, the causative agent may well be sexually transmitted. The effectiveness of antimicrobial agents in this condition remains uncertain. Some patients benefit from a 4- to 6-week course of treatment with erythromycin, doxycycline,TMP-SMZ, or a fluoroquinolone, but controlled trials are lacking.

PROSTATODYNIA

Patients who have symptoms and signs of prostatitis but who have no evidence of prostatic inflammation (normal leukocyte counts) and negative urine cultures are classified as having prostatodynia. Despite their symptoms, these patients most likely do not have prostatic infection and should not be given antimicrobial agents.

(Bibliography omitted in Palm version)

Back to Table of Contents

281. URINARY TRACT OBSTRUCTION - *Julian L. Seifter*, *Barry M. Brenner*

Obstruction to the flow of urine, with attendant stasis and elevation in urinary tract pressure, impairs renal and urinary conduit functions and is a common cause of acute and chronic renal failure. With early relief of obstruction, the defects in function usually disappear completely. However, chronic obstruction may produce permanent loss of renal mass (renal atrophy) and excretory capability, as well as enhanced susceptibility to local infection and stone formation. Early diagnosis and prompt therapy are therefore essential to minimize the otherwise devastating effects of obstruction on kidney structure and function.

ETIOLOGY

Obstruction to urine flow can result from *intrinsic* or *extrinsic mechanical blockade* as well as from *functional defects* not associated with fixed occlusion of the urinary drainage system. Mechanical obstruction can occur at any level of the urinary tract, from the renal calyces to the external urethral meatus. Normal points of narrowing, such as the ureteropelvic and ureterovesical junctions, bladder neck, and urethral meatus, are common sites of obstruction. When blockage is above the level of the bladder, unilateral dilatation of the ureter (*hydroureter*) and renal pyelocalyceal system (*hydronephrosis*) occur; lesions at or below the level of the bladder cause bilateral involvement.

Common forms of obstruction are listed in Table 281-1. In childhood, *congenital malformations*, including marked narrowing of the ureteropelvic junction, anomalous (retrocaval) location of the ureter, and posterior urethral valves, predominate. The latter defect is the most common cause of bilateral hydronephrosis in boys. Children may also have bladder dysfunction secondary to congenital urethral stricture, urethral meatal stenosis, or bladder neck obstruction. In adults, urinary tract obstruction is due mainly to *acquired defects*. Pelvic tumors, calculi, and urethral stricture predominate. Ligation of, or injury to, the ureter during pelvic or colonic surgery can lead to hydronephrosis which, if unilateral, may remain relatively silent and undetected. *Schistosoma haematobium* and genitourinary tuberculosis are infectious causes of ureteral obstruction. Obstructive uropathy may also result from extrinsic neoplastic (carcinoma of cervix or colon, retroperitoneal lymphoma) or inflammatory disorders. One such inflammatory disorder is retroperitoneal fibrosis, a process of unknown cause seen most commonly in middle-aged men, which occasionally leads to bilateral ureteral obstruction. Retroperitoneal fibrosis must be distinguished from other retroperitoneal causes of ureteral obstruction, particularly lymphomas and pelvic neoplasms.

Functional impairment of urine flow usually results from disorders that involve both the ureter and bladder. Common functional lesions include neurogenic bladder, often with adynamic ureter, and vesicoureteral reflux. Reflux of urine from bladder to ureter(s) is more common in children than in adults and may result in severe unilateral or bilateral hydroureter and hydronephrosis. Abnormal insertion of the ureter into the bladder is the most common cause of vesicoureteral reflux in children. Reflux in the absence of urinary tract infection or bladder neck obstruction usually does not lead to renal parenchymal damage and often resolves spontaneously as the child matures. Surgical reinsertion of the ureter into the bladder is indicated if reflux is severe and unlikely to improve spontaneously, if renal function deteriorates, or if urinary tract infections recur despite

chronic antimicrobial therapy. Hydronephrosis, usually more marked on the right than on the left, is common in pregnancy, due both to ureteral compression by the enlarged uterus and to functional effects of progesterone.

CLINICAL FEATURES

The pathophysiology and clinical features of urinary tract obstruction are summarized inTable 281-2. *Pain* is the symptom that most commonly provokes the need for medical attention. The pain of urinary tract obstruction is due to distention of the collecting system or renal capsule. The severity of the pain is influenced more by the rate at which distention develops than by the degree of distention. Acute supravesical obstruction, as from a stone lodged in a ureter (Chap. 279), is associated with excruciatingly severe pain, usually called *renal colic*. This pain is relatively steady and continuous, with little fluctuation in intensity, and often radiates to the lower abdomen, testes, or labia. By contrast, more insidious causes of obstruction, such as chronic narrowing of the ureteropelvic junction, may produce little or no pain yet result in total destruction of the affected kidney. Flank pain that occurs only with micturition is pathognomonic of vesicoureteral reflux.

Azotemia develops in urinary tract obstruction when overall excretory function is impaired. This may occur in the setting of bladder outlet obstruction, bilateral renal pelvic or ureteric obstruction, or unilateral disease in a patient with a solitary functioning kidney. Complete bilateral obstruction should be suspected when acute renal failure is accompanied by anuria. Any patient with renal failure otherwise unexplained or with a history of nephrolithiasis, hematuria, diabetes mellitus, prostatic enlargement, pelvic surgery, trauma, or tumor should be evaluated for urinary tract obstruction.

In the acute setting, bilateral obstruction may result in sodium and water retention that may mimic prerenal azotemia. However, with more prolonged obstruction, symptoms of *polyuria* and *nocturia* commonly accompany partial urinary tract obstruction and result from impaired renal concentrating ability. This defect usually does not improve with administration of vasopressin and is therefore a form of acquired nephrogenic diabetes insipidus. Disturbances in sodium chloride transport in the ascending limb of Henle and, in azotemic patients, the osmotic (urea) diuresis per nephron lead to decreased medullary hypertonicity and hence a concentrating defect. Partial obstruction, therefore, may be associated with increased rather than decreased urine output. Indeed, wide fluctuations in urine output in a patient with azotemia should always raise the possibility of intermittent or partial urinary tract obstruction. If fluid intake is inadequate, severe dehydration and hypernatremia may develop. Hesitancy and straining to initiate the urinary stream, postvoid dribbling, urinary frequency, and (overflow) incontinence are common with obstruction at or below the level of the bladder.

In addition to loss of urinary concentrating ability and azotemia, partial bilateral urinary tract obstruction often results in other derangements of renal function, including *acquired distal renal tubular acidosis*, *hyperkalemia*, and *renal salt wasting*. These defects in tubule function are often accompanied by renal tubulointerstitial damage. Morphologic abnormalities appear early in the course of obstruction; initially the interstitium becomes edematous and infiltrated with mononuclear inflammatory cells. With continued obstruction, the interstitium becomes fibrotic; scarring and atrophy of the

papillae and medulla occur and precede these processes in the cortex.

The possibility of urinary tract obstruction must always be considered in patients with urinary tract infections or urolithiasis. Urinary stasis encourages the growth of organisms as well as the formation of crystals, especially magnesium ammonium phosphate (struvite). *Hypertension* is frequent in acute and subacute unilateral obstruction and is usually a consequence of increased release of renin by the involved kidney. Chronic unilateral or bilateral hydronephrosis, in the presence of extracellular volume expansion or other renal disease, may result in significant hypertension. *Erythrocytosis*, an infrequent complication of obstructive uropathy, is probably secondary to increased erythropoietin production by the obstructed kidney.

DIAGNOSIS

A history of difficulty in voiding, pain, infection, or changes in urinary volume is common. Evidence for distention of the kidney or urinary bladder can often be obtained by palpation and percussion of the abdomen. A careful rectal examination may reveal enlargement or nodularity of the prostate, abnormal rectal sphincter tone, or a rectal or pelvic mass. The penis should be inspected for evidence of meatal stenosis or phimosis. In the female, vaginal, uterine, and rectal lesions responsible for urinary tract obstruction are usually revealed by inspection and palpation.

Urinalysis and examination of the urine sediment may reveal hematuria, pyuria, and bacteriuria. Often, however, the urine sediment is normal, even when obstruction leads to marked azotemia and extensive structural damage. An abdominal scout film should be obtained to evaluate the possibility of nephrocalcinosis or a radiopaque stone at any level of the urinary collecting system. As indicated inFig. 281-1, if urinary tract obstruction is suspected, a bladder catheter should be inserted. If diuresis does not follow, then abdominal ultrasonography should be performed to evaluate renal and bladder size, as well as pyelocalyceal contour. Ultrasonography is approximately 90% specific and sensitive for detection of hydronephrosis. False-positive results are associated with diuresis, renal cysts, or presence of an extrarenal pelvis, a normal congenital variant. Hydronephrosis may be absent on ultrasound when obstruction is associated with volume contraction, staghorn calculi, retroperitoneal fibrosis, or infiltrative renal disease.

In some cases, the intravenous urogram may define the site of obstruction. In the presence of obstruction, the appearance time of the nephrogram is often delayed. Eventually, however, the renal image becomes more dense than normal because of slow tubular fluid flow rate, which results in enhanced water reabsorption by the nephrons and greater concentration of contrast medium within tubules. The kidney involved by an acute obstructive process is usually slightly enlarged, and there is dilatation of the calyces, renal pelvis, and ureter above the obstruction. The ureter, however, is not tortuous, as is the case when the obstruction is chronic. In comparison with the nephrogram, the urogram may be extremely faint, especially if the dilated renal pelvis is voluminous, causing dilution of the contrast medium. The radiographic study should be continued until the site of obstruction is determined or the contrast medium is excreted. Radionuclide scans define less anatomic detail than intravenous urography and, like the urogram, are of limited value when renal function is poor. Nonetheless,

such scans are sensitive for the detection of obstruction and provide a substitute test in some patients at high risk for reaction to intravenous contrast.

To facilitate visualization of a suspected lesion in a ureter or renal pelvis, *retrograde* or *antegrade urography* should be attempted. These diagnostic studies may be preferable to the intravenous urogram in the azotemic patient, in whom poor excretory function precludes adequate visualization of the collecting system. Furthermore, intravenous urography carries the risk of contrast-induced acute renal failure in patients with proteinuria, renal insufficiency, diabetes mellitus, or multiple myeloma, particularly when performed under conditions of dehydration. The retrograde approach involves catheterization of the involved ureter under cystoscopic control, while the antegrade technique necessitates placement of a catheter into the renal pelvis via a needle inserted percutaneously under ultrasonic or fluoroscopic guidance. While the antegrade approach carries the added advantage of providing immediate decompression of a unilateral obstructing lesion, many urologists initially attempt the retrograde approach and resort to the antegrade method only when attempts at retrograde catheterization are unsuccessful or when cystoscopy or general anesthesia is contraindicated.

Patients suspected of having intermittent ureteropelvic obstruction (whether functional or mechanical) should have radiologic evaluation while they are in pain, since a normal urogram is commonly seen during asymptomatic periods. Hydration often helps to provoke a symptomatic attack. Voiding cystourethrography is of great value in the diagnosis of vesicoureteral reflux and bladder neck and urethral obstructions. Patients with obstruction at or below the level of the bladder exhibit thickening, trabeculation, and diverticula of the bladder wall. Postvoiding films reveal residual urine. If these radiographic studies fail to provide adequate information for diagnosis, endoscopic visualization by the urologist often permits precise identification of lesions involving the urethra, prostate, bladder, and ureteral orifices.

Computed tomography is useful in the diagnosis of specific intraabdominal and retroperitoneal causes of obstruction but is less practical as an initial test to establish the presence of obstruction. Magnetic resonance imaging may also be useful in the identification of specific obstructive causes.

TREATMENT

An individual with any form of urinary tract obstruction complicated by infection requires relief of obstruction as soon as possible to prevent development of generalized sepsis and progressive renal damage. On a temporary basis, depending on the site of obstruction, drainage is often satisfactorily achieved by nephrostomy; ureterostomy; or ureteral, urethral, or suprapubic catheterization. The patient with acute urinary tract infection and obstruction should be given appropriate antibiotics based on in vitro bacterial sensitivity and ability of the drug to concentrate in the kidney and urine. Treatment may be required for 3 to 4 weeks. Chronic or recurrent infections in an obstructed kidney with poor intrinsic function may necessitate nephrectomy. When infection is not present, immediate surgery often is not required, even in the presence of complete obstruction and anuria (because of the availability of dialysis), at least until acid-base, fluid and electrolyte, and cardiovascular status are restored to normal. Nevertheless, the site of obstruction should be ascertained as soon as feasible, in part

because of the possibility that sepsis may occur and necessitate prompt urologic intervention. Elective relief of obstruction is usually recommended in patients with urinary retention, recurrent urinary tract infections, persistent pain, or progressive loss of renal function. Infrequently, mechanical obstruction can be alleviated by nonsurgical means, as with radiation therapy for retroperitoneal lymphoma. Likewise, functional obstruction secondary to neurogenic bladder may be decreased with the combination of frequent voiding and cholinergic drugs.*The approach to obstruction secondary to renal stones is discussed in Chap. 279.*

PROGNOSIS

With relief of obstruction, the prognosis regarding return of renal function depends largely on whether irreversible renal damage has occurred. When obstruction is not relieved, the course will depend mainly on whether the obstruction is complete or incomplete, bilateral or unilateral, and whether urinary tract infection is also present. Complete obstruction with infection can lead to total destruction of the kidney within days. In dogs, relief of complete obstruction of 1 and 2 weeks' duration restores glomerular filtration rate to 60 and 30% of normal, respectively; after 8 weeks of obstruction, recovery does not occur. Nevertheless, in the absence of definitive evidence of irreversibility, every effort should be made to decompress in the hope of restoring renal function at least partially. A renal radionuclide scan, performed after a prolonged period of decompression, may be used to predict reversible renal function.

POSTOBSTRUCTIVE DIURESIS

Relief of bilateral, but not unilateral, complete urinary tract obstruction commonly leads to a postobstructive diuresis, characterized by polyuria, which may be massive. The urine is usually hypotonic and may contain large amounts of sodium chloride, potassium, and magnesium. The natriuresis is due, at least in part, to the excretion of retained urea, which acts as a poorly reabsorbable solute and diminishes salt and water reabsorption in the tubules (osmotic diuresis). The increase in intratubular pressure very likely also contributes to the impairment in net sodium chloride reabsorption, especially in the terminal nephron segments. Natriuretic factors (other than urea) may also accumulate during uremia induced by obstruction and depress salt and water reabsorption when urine flow is reestablished. In the majority of patients this diuresis is physiologic, resulting in the *appropriate* excretion of the excesses of salt and water retained during the period of obstruction. When extracellular volume and composition return to normal, the diuresis usually abates spontaneously. Therefore, replacement of urinary losses should serve only to prevent hypovolemia, hypotension, or disturbances in serum electrolyte concentrations. Occasionally, iatrogenic expansion of extracellular volume, secondary to administration of excessive quantities of intravenous fluids, is responsible for, or sustains, the diuresis observed in the postobstructive period. Replacement of no more than two-thirds of urinary volume losses per day is usually effective in avoiding this complication. The loss of electrolyte-free water with urea may result in hypernatremia. Serum and urine sodium and osmolal concentrations should guide the use of appropriate intravenous replacement. Often replacement with 0.45% saline is required. In a rare patient, relief of obstruction may be followed by urinary salt and water losses severe enough to provoke profound dehydration and vascular collapse. In these patients, an intrinsic defect in tubule reabsorptive function is probably

responsible for the marked diuresis. Appropriate therapy in such patients includes intravenous administration of salt-containing solutions to replace sodium and volume deficits.

(Bibliography omitted in Palm version)

Back to Table of Contents

PART ELEVEN -DISORDERS OF THE GASTROINTESTINAL SYSTEM

SECTION 1 -DISORDERS OF THE ALIMENTARY TRACT

282. APPROACH TO THE PATIENT WITH GASTROINTESTINAL DISEASE - *Daniel K. Podolsky*, *Kurt J. Isselbacher*

BIOLOGIC CONSIDERATIONS

The mucosal surface of the gastrointestinal (GI) tract is composed of a highly dynamic population of epithelial cells that are specialized for transmembrane absorption and secretion. These secretory and absorptive abilities facilitate digestion and nutrient uptake, which must be accomplished while keeping out potentially harmful pathogens and mutagens in the lumen. The barrier function is accomplished through both the physical integrity of the mucosal surface and the extensive population of resident immune cells.

The intestinal lymphoid system reflects a balance between dampening immune reactivity at the mucosal surface to prevent the constant and unrestrained activation of immune and inflammatory processes and immune response amplification in the underlying lamina propria and submucosa ready to respond when surface defenses have been breached. Derangements in the balance of suppression and stimulation predispose theGI tract to numerous inflammatory conditions.

The epithelial cells of the mucosal surface turn over very rapidly; the entire surface is regenerated every 24 to 72 h. This rapid turnover may permit rapid recovery of function following an acute insult and protect the cells against many mutagens in the lumen. Indeed, the small intestine rarely develops epithelial cancers. The slower turnover of colonic epithelium, the slower movement of the luminal contents, and differences in the mutagens in the luminal contents appear to foster colon cancers. Another fundamental feature of theGImucosa is the spatial segregation of the proliferative compartment from the terminally differentiated cells, especially in the small intestine, where a gradient of differentiation exists from the depths of the crypts of Lieberkuhn to the villus tip. This organization has a strong effect on the histology and pathophysiology of many mucosal disorders, such as celiac sprue.

Diseases of theGI tract produce clinical consequences through physical disruption of the mucosal layer (e.g., blood loss, fluid loss, pathogenic invasion) or nutritional derangements caused by impaired digestion and nutrient absorption. Focal or localized disease processes are more likely to disrupt mucosa; diffuse processes are more likely to alter absorption.

While the essential roles of theGItract -- the absorption of nutrients and the excretion of the products -- are accomplished in large part at the luminal surface, GI function also depends on the coordinated propulsion of food through the lumen by smooth-muscle contraction. The local and distant neural and endocrine factors that contribute to the regulation of intestinal motility are complex. Disruption of normal motility is common, with alterations in frequency of bowel movements, abdominal distention, abdominal pain, and nausea, individually or in varying combinations (so-called functional bowel

complaints), affecting as many as 15% of adults. Such symptoms may result from dysmotility related to the *direct* effects of an obstructing lesion or to the *indirect* actions of substances released by a primary mucosal disorder (e.g., inflammatory mediators such as arachidonic acid metabolites that also affect smooth-muscle activity).

The spectrum of diseases affecting theGI tract and their clinical manifestations are related to the component organ(s) involved (Table 282-1). Thus, esophageal disorders manifest themselves mainly through their effects on swallowing; gastric disorders are dominated by features relating to acid secretion; and diseases of the small and large intestine demonstrate disruption of nutrition and alterations of bowel movements. The GI tract may also be affected by systemic disorders, including vascular, inflammatory, infectious, and neoplastic conditions leading to focal or diffuse structural lesions. Metabolic and endocrine abnormalities as well as some drugs can disrupt normal bowel motility. When no structural lesion can be identified to explain GI symptoms, the disorder is termed *functional*.Table 282-2 summarizes criteria that may be used to distinguish functional from organic or structural diseases of the GI tract.

CLINICAL CONSIDERATIONS

History A thorough clinical history is essential in directing the clinician's attention to appropriate diagnostic considerations in the patient withGIsymptoms (Table 282-1). The most common complaints include pain and alterations in bowel habit, especially diarrhea or constipation. *Abdominal pain* is the most frequent and variable complaint and may reflect a broad spectrum of problems, from self-limited to urgent (Chap. 14). The intensity should be assessed and an initial distinction should be made between pain of acute onset and more chronic discomfort. Pain of abrupt onset more often reflects serious illness requiring urgent intervention, while a history of chronic discomfort is most often related to an indolent disorder. Dyspepsia, an ill-defined upper abdominal discomfort, is especially common and is often accompanied by varying degrees of nausea, bloating, and distention. Dyspepsia may be associated with peptic ulceration, but non-ulcer dyspepsia (NUD) is more common. A change in the pattern or character of pain may signify disease progression. Ascertaining the location of the pain (upper or lower, localized or diffuse), its character (sharp, burning, cramping), and its relationship to meals will often provide clues into the most important diagnostic considerations. Discomfort while the patient is eating suggests an esophageal disorder. Pain occurring shortly after the meal may signify biliary tract disease or abdominal angina; pain 30 to 90 min later is typical of peptic disease. Pain that is not affected by eating suggests a process outside of the bowel lumen, such as abscess, peritonitis, pancreatitis, and some malignancies. Conversely, factors that relieve the symptom are also helpful. For example, eating or antacid use typically relieves pain in peptic ulcer disease or gastritis. A relationship to bowel movement, especially together with an altered bowel habit, should focus attention on a disorder of the small or large bowel, such as inflammatory bowel disease.

Alterations in bowel habit can result from either disruption of normal intestinal motility or significant structural pathology. The temporal evolution of the change, the nature of the alteration, and the presence of other constitutional symptoms such as weight loss, fever, or anorexia are important. Temporary variation in bowel habit in association with some life stress and in the absence of signs of systemic illness suggests the common "irritable

bowel syndrome," especially when the alteration varies between diarrhea and constipation. Small, pellet-like stools associated with symptoms of dyspepsia (bloating, nausea, and "gas") are common. This diagnosis can essentially be made on the basis of a thorough history and physical examination and very limited laboratory testing, to exclude structural disease.

Constipation is a common complaint and may reflect an obstructing process but is more often due to impaired motility; though often functional in nature, drugs (e.g., anticholinergics), neurologic processes (e.g., Hirschsprung's disease), or smooth-muscle diseases (e.g., scleroderma) may cause decreased motility. The history and physical examination may provide evidence of a more generalized disorder such as hypothyroidism or depression. Pain associated with constipation may suggest an anal or perianal process with stool retention. The history may clarify that "constipation" actually reflects more an unrealized expectation of regularity than significant pathology. In contrast, progressively worsening constipation and weight loss in an adult with previously regular habits suggests the possible presence of an underlying obstructing process, particularly malignancy.

Although *diarrhea* refers to an increased frequency of movements, patients often use the term to describe loose or watery stools. If diarrhea is described, the daily average number of stools, their consistency, their pattern, and the presence of blood should be defined. The occurrence of nocturnal or true bloody diarrhea almost always reflects structural rather than functional bowel disease. A pungent stool odor or the presence of undigested meat in the movement is suggestive of pancreatic insufficiency. An alteration in color can be seen in cholestasis or steatorrhea (light-colored) or hemorrhage (melenic to maroon or bright red). Mucus in the movement is usually a sign of a functional bowel syndrome, while pus suggests infectious or inflammatory disease. Less common but more dramatic are the symptoms of acute GI bleeding, including hematemesis, melena, and hematochezia, which usually lead to prompt seeking of medical attention but should always be enquired after by the clinician.

In the evaluation of male patients, especially those with diarrhea or dysphagia, a tactful inquiry into sexual activity is essential. Homosexual males are at increased risk for a large variety of GI disorders as well as AIDS, which may first manifest itself with GI symptoms. AIDS patients are susceptible to a wide range of infections and neoplastic disorders of the GI tract, liver, and biliary tract (Chap. 308).

Finally, careful attention must be given to a general medical history with an emphasis on present or past use of medications or nonprescription drugs. Thyroid and other metabolic disorders, especially those affecting calcium metabolism, can cause a variety of GI symptoms. Unless asked, patients may forget to mention that they take aspirin almost daily for headache, and this may account for occult blood found in the stool. The use of daily laxatives may explain chronic diarrhea.

Physical Examination, Endoscopy, and Radiology All of the cardinal methods of examination are helpful in evaluating the patient with GI symptoms (Table 282-1). *Inspection* may disclose signs of cholestasis or nutritional deficiencies. Examination of the abdomen for an abnormal contour or inspection of the perianal region may reveal signs of a mass or a draining fistula. *Auscultation* may elicit a succussion splash in

patients with symptoms of gastric outlet obstruction. The absence of bowel sounds or an alteration in pitch can lead to recognition of an evolving ileus or an obstructing process. A bruit may be noted when symptoms of ischemic bowel disease are present. Careful *palpation* of the abdomen is especially important in detecting tenderness and masses, which can lead to the recognition of cholecystitis, Crohn's disease, periappendiceal abscess, and many other disorders. Findings on abdominal palpation will often be complemented by *percussion*, which is essential to assessing liver and spleen size.

Elicitation of *rebound tenderness*, either direct or referred, after abrupt removal of the examining hand provides an important clue to localized or more generalized peritonitis, which may suggest abdominal emergencies, such as a perforated viscus, intraabdominal abscess, or bowel infarction. Typically, the patient will remain immobile to avoid the accentuation of pain that may follow even slight movement or jarring of the abdomen. By contrast, patients with severe pain deriving from visceral disease, such as intestinal ischemia, are sometimes frantic to find a comfortable position. In these disorders, the absence of findings on palpation may be in striking contrast to the evident distress of the patient. Only when the process progresses to tissue destruction (e.g., intestinal infarction) and secondary peritonitis will the abdominal examination prove remarkable, often in concert with striking signs of systemic illness, including hemodynamic instability.

In addition to the examination of the abdomen, a digital rectal examination is also essential. In the patient with complaints of stool incontinence, the integrity of the sphincter can be assessed. Masses intrinsic to the rectum as well as abnormalities in the pelvis or the pouch of Douglas may only be detected by this examination. The presence of frank or occult blood in the stool is always important diagnostic information. Sigmoidoscopy should be viewed as a routine part of the physical examination in the patient with diarrhea, constipation, or frank or occult fecal blood. Sigmoidoscopy performed with either a rigid or a flexible fiberoptic instrument allows for direct inspection of the rectosigmoid mucosa, permitting the detection of cancers and polyps in this lower bowel segment that could be missed by barium x-rays. Inflammatory changes of the mucosa can help identify patients with infectious dysentery or other forms of colitis. Edema, granularity, diffuse friability (easily induced mucosal bleeding), and superficial ulcerations are characteristic of ulcerative colitis. Fresh stool samples for microbiologic studies and superficial mucosal biopsies obtained at the time of sigmoidoscopy can also yield crucial diagnostic information. The presence of polyps is an indication for colonoscopy (Chap. 283).

Many upper and lowerGI tract disorders are accessible to inspection via fiberoptic instruments. As a result, endoscopy has supplanted conventional contrast x-ray studies for many clinical problems, both because of its heightened precision for diagnosis and the opportunity in many instances to accomplish meaningful therapeutic intervention. However, it should be emphasized that *no procedure should be considered routine* and used indiscriminately; there must be a rational basis for its use in the individual patient. These techniques are discussed in detail inChap. 283. Upper GI endoscopy permits evaluation of the esophagus, stomach, duodenum, and, with specially designed instruments, proximal jejunum. Side-viewing scopes permit inspection and cannulation of the ampulla of Vater, facilitating retrograde cholangiopancreatography. Evaluation of some patients will be further benefited by endoscopic ultrasound (US), which can

delineate submucosal mass lesions and abnormalities in the pancreas. The colonoscope can be used to visualize the entire colon and often the terminal ileum, resulting in more accurate diagnosis of inflammatory bowel disease and mass lesions. Colonic polyps can almost always be removed at the time of their initial identification.

Endoscopic techniques are relatively precise in defining many problems, but the limitations of these tools, as well as the continued advantages of x-ray studies in some situations, should be recognized. Endoscopic tools are not useful in assessingGImotility, which may be assessed more accurately by barium studies. In addition, the small intestine remains largely inaccessible to fiberoptic instruments. In hospitals where endoscopy is not feasible, the upper GI series and barium enema remain good diagnostic modalities to evaluate the upper and lower GI tract, especially when air-contrast techniques are employed. However, they should generally be avoided in patients with GI bleeding or suspected bowel obstruction. In addition, the cathartics used to prepare the bowel may markedly worsen the condition of a patient with obstructing lesions or colitis.

Although endoscopy has obviated the need for many conventionalGIx-rays, other radiologic imaging modalities, includingUS, computed tomography (CT), and magnetic resonance imaging (MRI), have assumed a larger role in patients with GI symptoms. Both US and CT are useful in the delineation of abdominal masses. CT, though more expensive, is often more effective in the evaluation of the lower abdomen, where inflammatory masses in patients with Crohn's disease or complications of diverticular disease may be accurately imaged. However, US is an effective and less expensive tool for the evaluation of the right upper quadrant, including the gall bladder and biliary tract. MRI may give exquisitely accurate information on the anatomic extent of invasive rectal cancers and blood flow in patients with vascular disorders, but the full range of its uses in GI disorders remains to be delineated. More sophisticated CT and MRI equipment can actually permit the performance of digital angiography without the invasive catheterization necessary in conventional visceral angiography. CT "virtual colonoscopy," a nonendoscopic method of visualizing the colon, is developing rapidly.

Radionuclide scans can be used to localize a site of bleeding in theGItract. Radiolabeled technetium can detect a Meckel's diverticulum, which is an occasional source of bleeding.

DIAGNOSTIC APPROACHES (Table 282-1)

Abdominal Pain Determining the cause of abdominal pain is frequently a clinical challenge (Chap. 14). Differential diagnostic considerations may encompass diseases extrinsic to theGItract, such as disorders of the genitourinary tract (e.g., pelvic inflammatory disease) and the peritoneum. The initial goal is to distinguish between an urgent problem and a nonacute disorder. Initial clinical impressions based on the history and physical examination can be further refined through routine laboratory tests such as a complete blood count and differential as well as plain films of the abdomen. Specific features will dictate the appropriateness of urgentUS orCTexamination or the need to proceed promptly with surgery. In the patient with a long-standing and relatively stable problem, diagnostic evaluation can be more deliberate. A functional basis for the complaint may be established on the strength of the history and physical examination

alone. Radiologic contrast studies, other imaging modalities (e.g., US, CT), or endoscopic examination may be appropriate. If these approaches do not determine the cause of the patient's symptoms, more unusual causes of abdominal pain such as acute intermittent porphyria may have to be excluded through specific urine or blood tests (Chap. 346).

Problems of Swallowing Dysphagia nearly always signifies the presence of structural pathology. The approach should be as follows:

1. *Thorough determination of the nature of dysphagia*. Is the difficulty primarily in swallowing liquids, solids, or both? The location of the difficulty from the patient's perspective and presence or absence of accompanying *odynophagia* (pain on swallowing) are important to ascertain. These historic clues are complemented by careful visual and neurologic examination of the oropharynx.

2. *Routine esophageal x-rays* in the upright and lateral or Trendelenburg position. The horizontal views are essential for demonstration of the swallowing mechanism, unaided by gravity, and of the esophagogastric junction. For details of the pharyngoesophageal area, cineradiography is necessary because of the rapidity with which the contrast medium passes through. Hiatus hernia is extremely common (in 15 to 35% of persons over 50) and is often asymptomatic. Careful attention is usually needed to detect lower esophageal rings or webs, which may be visible as indentations in the barium column only from a limited angle.

3. *Esophagoscopy*. This procedure is desirable to biopsy masses or abnormal mucosa and to obtain washings for exfoliative cytologic study. The diagnoses of peptic esophagitis and Barrett's esophagus are made endoscopically. Endoscopy is the most sensitive technique for identifying esophageal or gastric varices, although they are seldom important in the absence of hemorrhage. Endoscopic instruments with aUS probe at the tip (endoscopic ultrasound) are useful diagnostic and staging tools for certain problems of the esophagus (and other sites of theGItract).

4. *Manometric studies* of the upper esophagus, particularly in conjunction with cineradiography. This procedure offers the best means of differentiating among disorders originating in the central nervous system, primary pharyngeal muscular disease, and cricopharyngeal dystonia. Manometry of the lower esophagus is useful in the diagnosis of diffuse esophageal spasm, achalasia, and infiltrative diseases that alter esophageal motility.

5. *24-Hour monitoring of esophageal pH* may be used to document esophageal reflux.

Peptic or Digestive Disorders The approaches to these disorders include the following:

1. *Insertion of a nasogastric tube.* This approach is used to establish whether significant gastric retention (more than 75 mL of gastric contents in the fasting state) exists and whether acid, bile, blood, or other materials are present. If pyloric obstruction or gastric atony is present, the tube is used to maintain suction while the patient's electrolyte and fluid balance is restored to normal; the stomach is kept as clean as possible so that

diagnostic investigation may be carried out.

2. *Upper gastrointestinal endoscopy* (Chap. 283). This procedure is most helpful in assessing the mucosa in gastritis or, together with biopsy and brushings for cytology, in differentiating between peptic and neoplastic ulcerating lesions. It may identify a specific bleeding site in clinical situations where several potential bleeding sites could exist, as in the patient with portal hypertension. In addition, it may be possible to cauterize or otherwise intervene to control hemorrhage via the endoscope (e.g., by injections of vasoconstricting agents such as epinephrine). *Helicobacter pylori* is a frequent cause of gastritis in patients with peptic ulceration and non-ulcer dyspepsia. Although *H. pylori* infection can be confirmed by endoscopy and biopsy, the diagnosis is more commonly made by breath and serologic tests (Chap. 285). Endoscopy is the diagnostic method of choice in the setting of upperGIbleeding (Chap. 44). Endoscopy can detect a number of potential sources of upper GI bleeding that are often missed by x-ray studies (e.g., erosive gastritis, Mallory-Weiss tear). Gastroscopy is particularly helpful in inspecting the postoperative stomach, especially in detecting stomal ulceration or so-called alkaline reflux gastritis. The first and second portions of the duodenum can also be routinely examined, and important information about ulcers and other lesions can be obtained. Radiologic studies may be useful when endoscopy is not readily available or in the assessment of suspected motility disorders (e.g., gastroparesis). In addition, radiologic examination may be preferred when there are contraindications to safe endoscopy.

3. *Gastric acid secretory studies.* Although not routinely necessary, these studies are useful in the diagnosis of the Zollinger-Ellison syndrome or atrophic gastritis and for determination of completeness of vagotomy. They should not be performed for the routine diagnosis of uncomplicated duodenal ulcer or to influence the choice of surgery for peptic ulcer.

Obstructive and Vascular Disorders of the Small Intestine (See alsoChaps. 289 and290) The plain x-ray film of the abdomen is the most important diagnostic adjunct to careful physical examination in patients with symptoms of obstruction. Patterns of dilation of individual loops of intestine may be characteristic, as in volvulus or acute pancreatitis; erect and decubitus views will often show fluid levels in the affected segments. Motility disorders of the small intestine (temporary ileus or chronic intestinal pseudoobstruction) may also present with obstructive symptoms and similar x-ray findings but must be managed medically without surgical intervention. Air under the diaphragm is diagnostic of a perforated viscus; air in the portal vein usually results from intestinal necrosis from mesenteric vascular occlusion. The diagnostic accuracy of the plain x-ray film in all types of intestinal obstruction is about 75%. In patients with symptoms of incomplete obstruction, the radiographic small-bowel series will often be diagnostic in defining the site and degree of obstruction. Infrequently, in this setting, all conventional x-ray studies are unremarkable. In such cases, the radiologist may perform a small-bowel enteroclysis study by passing a special tube into the proximal jejunum; the rapid instillation of barium through the tube will distend the intestine and often reveal subtle lesions missed by other tests.

Vascular diseases of the small intestine are among the most difficult diseases to diagnose. In chronic mesenteric ischemia, radiographic, endoscopic, and laboratory tests are usually normal. Early in the course of acute mesenteric ischemia, the plain film

of the abdomen may be unremarkable despite complaints of severe abdominal pain. In these settings, prompt mesenteric angiography is essential to confirm the diagnosis of vascular disease.

Inflammatory and Neoplastic Diseases of Small and Large Intestine Patients with these conditions are usually identified by history, physical examination, and careful examination of the stools for exudate and blood. Examination of fresh stool samples for common bacterial pathogens and parasites by laboratories skilled in these techniques is important in identifying or excluding infectious causes of diarrhea, particularly in the patient with colitis. Sigmoidoscopy is valuable in identifying mucosal and neoplastic lesions of the rectum and distal colon. The mucosal surface of the entire colon and terminal ileum can be examined directly and biopsied through the fiberoptic sigmoidoscope or colonoscope. The radiologic examination of the small intestine is highly reliable in identifying the prestenotic and stenotic lesions of Crohn's disease. In the colon, a single barium enema examination in a well-prepared patient has a diagnostic accuracy of 80 to 85%; the addition of air-contrast technique brings the accuracy up over 90%. Accuracy is greatly limited if the patient is poorly prepared for the examination. Colonoscopy may be preferable because of its greater accuracy and the fact that it enables the operator to remove any polyps that are encountered and to obtain preoperative tissue confirmation in the patient who probably has cancer.

Peroral biopsy of the small intestine (now most often accomplished during endoscopy) and forceps biopsy of the rectosigmoid are of considerable importance in revealing mucosal disease. Rectal biopsy is an excellent means of demonstrating amyloidosis, schistosomiasis, and amebiasis. Submucosal disease is not seen in these superficial biopsies. Hirschsprung's disease is diagnosed histologically by a deep surgical biopsy of the lower part of the rectum.

Malabsorption Syndromes Malabsorption may be suspected on the basis of history and physical examination and confirmed by examination of the stool. Radiologic examination is helpful to rule out local lesions and to suggest motor and secretory dysfunction, but it is rarely diagnostic unless an abnormal small-bowel mucosa or fistulas between the intestine and stomach are demonstrated.

Microscopic examination of a stool specimen stained with Sudan is a simple screening test for steatorrhea. Chemical analysis of 3-day stool collection for fat, with the patient on a standard diet, is used to establish the diagnosis of steatorrhea. The D-xylose absorption test is about 90% accurate in distinguishing mucosal disease from pancreatic insufficiency. Peroral biopsy of the small intestine via the endoscope or a specialized biopsy device is of value in the diagnosis of celiac disease, and it may show the less common infiltrations of the mucosa by amyloid or bacterial mucoproteins (Whipple's disease). Leakage of protein into the intestinal lumen may cause hypoproteinemia and can be demonstrated by the recovery in stools of the serum protein a_1-antitrypsin or intravenously administrated markers such as iodine- or chromium-labeled isotopes.***The tests useful in the diagnosis of malabsorption are discussed in Chap. 286.*

GIBleeding (See also Chap. 44) Acute bleeding in the GI tract is a common clinical problem. The history usually provides a reliable distinction between lower and upper tract sources. Once the patient with upper tract bleeding is hemodynamically stable, a

nasogastric tube is placed to confirm the site of blood loss and to empty the stomach. Endoscopy is then performed to define the cause and often to treat it. In patients with acute lower tract bleeding, sigmoidosopy may permit detection of distal sites of bleeding. Colonoscopy may also be of value, but visualization may be limited by active bleeding and poor bowel preparation. Barium studies should be avoided in the acute setting. They are usually nondiagnostic and the persistent contrast may interfere with interpretation of angiographic studies, which can often define a site of bleeding that is otherwise obscure. Radionuclide bleeding scan can locate the bleeding site. A Meckel's scan can be diagnostic when active bleeding arises distal to the duodenum in the absence of an identifiable source in the colon.

(Bibliography omitted in Palm version)

Back to Table of Contents

283. GASTROINTESTINAL ENDOSCOPY - *Mark Topazian*

Gastrointestinal endoscopy has been attempted for over 200 years, but the introduction of semi-rigid gastroscopes in the middle of the twentieth century marked the dawn of the modern endoscopic era. Since then, rapid advances in endoscopic technology have led to dramatic changes in the diagnosis and treatment of many digestive diseases. Innovative endoscopic devices and new endoscopic treatment modalities continue to expand the use of endoscopy in patient care.

Flexible endoscopes provide either an optical image (transmitted over fiberoptic bundles) or an electronic video image (generated by a charge-coupled device in the tip of the endoscope; see Color Atlas, Section III). Operator controls permit deflection of the endoscope tip; fiberoptic bundles bring light to the tip of the endoscope; and working channels allow washing, suctioning, and the passage of instruments. Progressive changes in the diameter and stiffness of endoscopes have improved the ease and patient tolerance of endoscopy.

ENDOSCOPIC PROCEDURES

Upper Endoscopy Upper endoscopy, also referred to as esophagogastroduodenoscopy (EGD), is performed by passing a flexible endoscope through the mouth into the esophagus, stomach, bulb, and second duodenum. The procedure is the best method of examining the upper gastrointestinal mucosa. While the upper gastrointestinal radiographic series has similar accuracy for diagnosis of duodenal ulcer, EGD is superior for detection of gastric ulcers and permits directed biopsy and endoscopic therapy, if needed. Topical pharyngeal anesthesia is used, and intravenous conscious sedation is given to most patients in the United States to ease the anxiety and discomfort of the procedure, although in many countries EGD is routinely performed without sedation. The recent development of ultrathin, 5-mm diameter endoscopes for transnasal, unsedated EGD may decrease the use of sedation for EGD in the United States, also decreasing the costs and risks of the procedure.

Colonoscopy Colonoscopy is performed by passing a flexible colonoscope through the anal canal into the rectum and colon. The cecum is reached in over 95% of cases, and the terminal ileum can often be examined. Colonoscopy is the "gold standard" for diagnosis of colonic mucosal disease. Barium enema is more accurate for evaluation of diverticula and for accurate measurement of colonic strictures, but colonoscopy has greater sensitivity for polyps and cancers. Colonoscopy is more uncomfortable thanEGD for most patients, and conscious sedation is usually given before colonoscopy in the United States, although a willing patient and a skilled examiner can complete the procedure without sedation in many cases.

Flexible Sigmoidoscopy Flexible sigmoidoscopy is similar to colonoscopy but visualizes only the rectum and a variable portion of the left colon, typically to 60 cm from the anal verge. This procedure causes abdominal cramping, but it is brief and is almost always performed without sedation. Flexible sigmoidoscopy is primarily used to screen asymptomatic, average-risk patients for colonic polyps and may also be used for evaluation of diarrhea and hematochezia.

Enteroscopy Enteroscopy is the relatively new field of small-bowel endoscopy. Two techniques are currently used. "Push" enteroscopy is performed with a long endoscope similar in design to an upper endoscope. The enteroscope is pushed down the small bowel with the help of a stiffening overtube that extends from the mouth to the duodenum. The mid-jejunum can often be reached; an instrument channel is present for biopsies or endoscopic therapy. "Sonde" enteroscopy uses a very thin, long, flexible endoscope with a weighted tip and no biopsy capability. The sonde enteroscope is passed through the nose, dragged to the duodenum by a standard endoscope, then slowly propelled forward by intestinal peristalsis for several hours. The cecum or distal ileum is reached in most cases. The small-bowel mucosa is examined during sonde enteroscope withdrawal, although parts of the mucosa may be missed when the endoscope is pulled back around turns. The major indication for these procedures is unexplained small-bowel bleeding.

Endoscopic Retrograde Cholangiopancreatography (ERCP) During ERCP, a side-viewing endoscope is passed through the mouth to the duodenum, the ampulla of Vater is identified and cannulated with a thin plastic catheter, and radiographic contrast material is injected into the bile duct and pancreatic duct under fluoroscopic guidance (Fig. 283-1). When indicated, the sphincter of Oddi can be opened using the technique of endoscopic sphincterotomy (Fig. 283-2). Stones can be retrieved from the ducts, and strictures of the ducts can be biopsied, dilated, and stented. ERCP is often performed for therapy but remains an important diagnostic tool, especially for bile duct stones.

Endoscopic Ultrasound (EUS) EUS utilizes high-frequency ultrasound transducers incorporated into the tip of a flexible endoscope. Ultrasound images are obtained of the gut wall and adjacent organs, vessels, and lymph nodes. By sacrificing depth of ultrasound penetration and bringing the ultrasound transducer close to the area of interest via endoscopy, very high resolution images are obtained. EUS provides the most accurate preoperative local staging of esophageal, pancreatic, and rectal malignancies, although it does not detect most distant metastases. Examples of EUS tumor staging are shown inFig. 283-3. EUS is also highly sensitive for diagnosis of bile duct stones, gallbladder disease, submucosal gastrointestinal lesions, and chronic pancreatitis. Fine-needle aspiration of masses and lymph nodes in the posterior mediastinum, abdomen, and pelvis can be performed under EUS guidance.

RISKS OF ENDOSCOPY

All endoscopic procedures carry some risk of bleeding and gastrointestinal perforation. These risks are quite low with diagnostic upper endoscopy and colonoscopy (<1:1000 procedures), although the risk is as high as 1:100 when therapeutic procedures such as polypectomy, control of hemorrhage, or stricture dilation are performed. Bleeding and perforation are rare with flexible sigmoidoscopy. The risks for diagnosticEUS are similar to the risks for diagnostic upper endoscopy.

Infectious complications are unusual with most endoscopic procedures. Stricture dilation, variceal sclerotherapy, andERCP for biliary obstruction all carry a higher incidence of postprocedure bacteremia, and prophylactic antibiotics may be indicated for these procedures in some patients (Table 283-1).

ERCPcarries additional risks. Pancreatitis occurs in about 5% of patients undergoing ERCP and is seen in up to 25% of patients with sphincter of Oddi dysfunction. Post-ERCP pancreatitis is usually mild and self-limited but may infrequently result in prolonged hospitalization, surgery, diabetes, or death. Bleeding occurs after 1% of endoscopic sphincterotomies. Ascending cholangitis, pseudocyst infection, and retroperitoneal perforation and abscess may all occur as a result of ERCP.

The conscious sedation administered during endoscopy may cause respiratory depression or allergic reactions. Percutaneous gastrostomy tube placement duringEGD is associated with a 10 to 15% incidence of complications, most often wound infections. Fasciitis, pneumonia, bleeding, and colonic injury may result from gastrostomy placement.

URGENT ENDOSCOPY

ACUTE GASTROINTESTINAL HEMORRHAGE

Endoscopy is an important diagnostic and therapeutic technique for patients with acute gastrointestinal hemorrhage. Although most gastrointestinal bleeding stops spontaneously, a minority of patients will have persistent or recurrent hemorrhage that may be life-threatening. Clinical predictors of rebleeding help identify patients most likely to benefit from urgent endoscopy and endoscopic, angiographic, or surgical hemostasis.

Initial Evaluation The initial evaluation of the bleeding patient focuses on the magnitude of hemorrhage as reflected by the postural vital signs, the frequency of hematemesis or melena, and (in some cases) findings on nasogastric lavage. The measured values of hematocrit and hemoglobin lag the clinical course and are not reliable gauges of the magnitude of acute bleeding. This initial evaluation, completed well before the bleeding source is confidently identified, guides immediate supportive care of the patient and helps determine the timing of endoscopy. The magnitude of the initial hemorrhage is probably the most important indication for urgent endoscopy, since a large initial bleed increases the likelihood of ongoing or recurrent bleeding. Patients with resting hypotension, repeated hematemesis, nasogastric aspirate that does not clear with repeated lavage, or those requiring blood transfusions should be considered for urgent endoscopy. In addition, patients with cirrhosis, coagulopathy, or respiratory or renal failure and those over 70 years of age are more likely to have significant rebleeding.

Bedside evaluation also suggests an upper or lower gastrointestinal source of bleeding in most patients. About 90% of patients with melena are bleeding proximal to the ligament of Treitz, and about 90% of patients with hematochezia are bleeding from the colon. It is important to note, however, that melena can result from bleeding in the small bowel or right colon, especially in older patients with slow colonic transit, so colonoscopy should be performed in patients with melena when upper endoscopy is unrevealing. Similarly, a minority of patients with massive hematochezia are bleeding from a duodenal ulcer, with rapid intestinal transit. Hence early upper endoscopy should be considered in patients with massive hematochezia.

Endoscopy should be performed after the patient has been resuscitated with

intravenous fluids and transfusions as necessary. Marked coagulopathy or thrombocytopenia is usually treated before endoscopy, since correction of these abnormalities may lead to resolution of bleeding, and techniques for endoscopic hemostasis are limited in such patients. Metabolic derangements should also be addressed. Tracheal intubation for airway protection should be considered before upper endoscopy in patients with repeated hematemesis and suspected variceal hemorrhage.

Most patients with impressive hematochezia can undergo colonoscopy after a rapid colonic purge with a polyethylene glycol solution; the preparation fluid is often administered via a nasogastric tube. In a minority of cases, persistent bleeding and recurrent hemodynamic instability prevent endoscopic visualization of the colonic mucosa, and other techniques (such as bleeding scans, angiography, or emergency subtotal colectomy) must be employed. Even in these cases, however, the anal and rectal mucosa should be visualized endoscopically early in the course, since bleeding lesions in or close to the anal canal are generally amenable to surgical transanal hemostatic techniques; and upper endoscopy should be performed to exclude duodenal ulcer.

Peptic Ulcer The endoscopic appearance of peptic ulcers provides useful prognostic information in patients with acute hemorrhage. When a platelet plug is seen protruding from a vessel wall in the base of an ulcer (a so-called sentinel clot or visible vessel), there is a 40% chance of major rebleeding from the ulcer. This finding often leads to local endoscopic therapy to decrease the rebleeding rate. A clean-based ulcer, on the other hand, is associated with low (3 to 5%) risk of rebleeding; patients with melena and a clean-based duodenal ulcer are often discharged to home from the emergency department or endoscopy suite if they are young, reliable, and otherwise healthy. Other findings have an intermediate risk of rebleeding: flat red or purple spots in the ulcer base have a 10% risk, and large adherent clots covering the ulcer base have a 20% risk. Occasionally, active spurting from an ulcer is seen (with>90% risk of ongoing bleeding). Examples of endoscopic stigmata of recent hemorrhage are shown inFig. 283-4.

Patients with a visible vessel or active bleeding are usually treated endoscopically, decreasing rebleeding rates by about half. Hemostatic techniques include "coaptive coagulation" of the vessel in the base of the ulcer, using a thermal probe that is pressed against the site of bleeding, or injection of epinephrine or sclerosant into and around the vessel.

Varices Two complementary strategies guide therapy of bleeding varices: local treatment of the bleeding vessel and treatment of underlying portal hypertension. Local therapies (including endoscopic sclerotherapy, endoscopic band ligation, and balloon tamponade with a Sengstaken-Blakemore tube) effectively control acute hemorrhage in most patients and are the mainstay of acute treatment, although therapies that decrease portal pressures (pharmacologic treatment, surgical shunts, or radiologically placed intrahepatic shunts) also play an important role.

Endoscopic band ligation is the preferred local therapy for bleeding esophageal varices. In this technique a varix is suctioned into a cap fitted on the end of the endoscope, and a rubber band is then released from the cap, ligating the varix. Acute hemorrhage can

be controlled in up to 90% of patients, and complications (such as sepsis, symptomatic esophageal ulceration, or esophageal stenosis) are uncommon. Endoscopic sclerotherapy is an older technique in which a sclerosing, thrombogenic solution is injected into or next to esophageal varices. Sclerotherapy also controls acute hemorrhage in most patients but has higher complication rates. These techniques are used when varices are actively bleeding during endoscopy or (more commonly) when varices are the only identifiable cause of acute hemorrhage.

After treatment of the acute hemorrhage, an elective course of endoscopic therapy can be undertaken with the goal of eradicating esophageal varices and preventing rebleeding months to years later. This chronic therapy is less successful, preventing long term rebleeding in about 50% of patients. Pharmacologic therapies that decrease portal pressure have similar efficacy, and the two modalities may be combined.

Gastric varices are less amenable to endoscopic therapy and are usually treated with a portal decompressive procedure (surgical portosystemic shunt or radiologic transjugular portosystemic shunt). Endoscopic therapy of gastric varices is usually reserved for actively bleeding varices or for patients with thrombosis of the portal venous system.

Dieulafoy's Lesion This lesion, also called *persistent caliber artery*, is a large-caliber arteriole that runs immediately beneath the gastrointestinal mucosa and bleeds through a pinpoint mucosal erosion. Dieulafoy's lesion is seen most commonly on the lesser curvature of the proximal stomach, causes impressive arterial hemorrhage, and is difficult to diagnose; it is often recognized only after repeated endoscopy for recurrent bleeding. Endoscopic therapy with a thermal probe usually controls acute bleeding and successfully ablates the underlying vessel once the bleeding site has been identified. Embolization or surgical oversewing are sometimes required.

Mallory-Weiss Tear A Mallory-Weiss tear is a linear mucosal rent near or across the gastroesophageal junction that is often associated with retching or vomiting. When the tear disrupts a submucosal arteriole, brisk hemorrhage may result. Endoscopy is the best method of diagnosis, and an actively bleeding tear can be treated endoscopically with coaptive coagulation using a thermal probe or by injection of dilute epinephrine. Since Mallory-Weiss tears only rarely rebleed, a sentinel clot in the base of the tear is usually not treated endoscopically.

Vascular Ectasias Vascular ectasias are flat mucosal vascular anomalies best diagnosed by endoscopy. They usually cause slow intestinal blood loss and have several characteristic distributions in the gastrointestinal tract. When limited to the cecum, where they occur as senile lesions, or the gastric antrum (gastric antral vascular ectasias, or "watermelon stomach"), ectasias are often responsive to local endoscopic ablative therapy. Patients with diffuse small-bowel vascular ectasias (associated with chronic renal failure and with hereditary hemorrhagic telangiectasia) often continue to bleed despite endoscopic treatment of accessible lesions and require systemic therapy.

Colonic Diverticula Diverticula form where nutrient arteries penetrate the muscular wall of the colon en route to the colonic mucosa. The artery found in the base of a diverticulum may bleed, causing painless and impressive hematochezia. Colonoscopy is indicated in patients with hematochezia and suspected diverticular hemorrhage, since

other causes of bleeding (such as vascular ectasias, colitis, and colonic malignancy) must be excluded. In addition, an actively bleeding diverticulum is occasionally seen and treated during colonoscopy.

GASTROINTESTINAL OBSTRUCTION AND PSEUDOOBSTRUCTION

Endoscopy is useful for evaluation and treatment of some forms of gastrointestinal obstruction. An important exception is small-bowel obstruction, which is generally not diagnosed by endoscopy or amenable to endoscopic therapy. Esophageal, gastroduodenal, and colonic obstruction or pseudoobstruction can all be diagnosed endoscopically and are often managed endoscopically as well.

Acute Esophageal Obstruction Esophageal obstruction by impacted food or an ingested foreign body is a potentially life-threatening event. Left untreated, the patient may develop esophageal ulceration, ischemia, and perforation. Patients with persistent esophageal obstruction often have hypersalivation and are usually unable to swallow water; endoscopy is generally the best initial test in such patients, since endoscopic removal of the obstructing material is usually possible, and the presence of an underlying esophageal stricture can often be determined. Radiographs of the chest and neck should be considered before endoscopy in patients with fever, obstruction for $^{3}24$ h, or ingestion of a sharp object such as a fishbone. Radiographic contrast studies interfere with subsequent endoscopy and are not advisable in patients with a clinical picture of persistent obstruction, unless an esophageal perforation is suspected. Occasionally, sublingual nifedipine or nitrates, or intravenous glucagon, may resolve an esophageal food impaction, but in most patients there is an underlying web, ring, or stricture and endoscopic removal of the obstructing food bolus is necessary.

Gastric Outlet Obstruction Obstruction of the gastric outlet is commonly caused by malignancy of the prepyloric gastric antrum or chronic peptic ulceration with stenosis of the pylorus. Patients vomit partially digested food many hours after eating. Gastric decompression with a nasogastric tube and subsequent lavage for removal of retained material is the first step in treatment. The diagnosis can then be confirmed with a saline load test, if desired. Endoscopy is useful for diagnosis and treatment. Patients with pyloric stenosis may be treated with endoscopic balloon dilation of the pylorus, and a course of endoscopic dilation results in long-term relief of symptoms in about 50% of patients. Malignant pyloric obstruction can be treated with endoscopically placed expandable stents if the patient is deemed a poor surgical candidate.

Colonic Obstruction and Pseudoobstruction These both present with abdominal distention and discomfort; tympany; and a dilated, air-filled colon on plain abdominal radiography. Both conditions may lead to colonic perforation if untreated. Acute colonic pseudoobstruction is a form of colonic ileus that is usually attributable to electrolyte disorders, narcotic and anticholinergic medications, immobility (as after surgery), and retroperitoneal hemorrhage or mass. Multiple causative factors are often present. Either colonoscopy or a water-soluble contrast enema may be used to look for an obstructing lesion and differentiate obstruction from pseudoobstruction. One of these diagnostic studies should be strongly considered if the patient does not have clear risk factors for pseudoobstruction, if radiographs do not show air in the rectum and sigmoid, or if the patient fails to improve when the underlying causes of pseudoobstruction have been

addressed. The risk of cecal perforation in pseudoobstruction rises when the cecal diameter exceeds 12 cm, and in such patients decompression of the colon may be achieved using intravenous neostigmine, colonoscopic decompression, or placement of a cecostomy tube. Most patients should receive a trial of conservative therapy (with correction of electrolyte disorders, removal of offending medications, and increased mobilization) before undergoing an invasive decompressive procedure.

Colonic obstruction is an indication for urgent surgery. In poor operative candidates or those with symptomatic partial obstruction from malignancy, a colonoscopically placed expandable stent can relieve obstruction and permit preparation of the bowel for elective surgery.

ACUTE BILIARY OBSTRUCTION

The steady, severe pain that occurs when a gallstone acutely obstructs the common bile duct often brings patients to a hospital. The diagnosis of a ductal stone is suspected when the patient is jaundiced or when serum liver tests or pancreatic enzyme levels are elevated, and it is confirmed by direct cholangiography (performed endoscopically, percutaneously, or during surgery).ERCP is currently the primary means of diagnosing and treating common bile duct stones in most hospitals in the United States.

Bile Duct Imaging While traditional noninvasive imaging tests such as ultrasound and biliary scintigraphy are not sufficiently accurate for reliable diagnosis of bile duct stones, newer imaging modalities such as spiral computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), andEUS are more accurate and have an emerging role in diagnosis. Examples of these modalities are shown in Fig. 283-5. During MRCP, images are obtained that demonstrate stagnant or slowly flowing fluid and subtract all other tissue. The resulting images of the right upper quadrant are strikingly similar to a direct cholangiogram, although with less resolution. MRCP can be performed rapidly without sedation and does not require any radiographic contrast. When an echo-endoscope is passed into the duodenum, detailed EUS views of the adjacent bile duct are readily obtained. While this procedure requires intravenous sedation, it has a very low incidence of complications, in contradistinction toERCP. Spiral CT has a sensitivity of 85% for diagnosis of bile duct stones, MRCP has a sensitivity of 85 to 95%, and EUS has a sensitivity of 88 to 98%. EUS is more accurate than ERCP in some hands.

The clinical role of these new imaging techniques is evolving. When a bile duct stone is highly likely and urgent treatment is required (as in a patient with jaundice and biliary sepsis),ERCP is the procedure of choice, since it remains the gold standard for diagnosis and provides immediate treatment. When a persistent bile duct stone is relatively unlikely (as in a patient with gallstone pancreatitis), less-invasive imaging techniques may supplant ERCP or intraoperative cholangiography.

Ascending Cholangitis Charcot's triad of jaundice, abdominal pain, and fever is present in about 70% of patients with ascending cholangitis and biliary sepsis. Initially, such patients are managed with fluid resuscitation and intravenous antibiotics. Abdominal ultrasound is often done early in the course, to look for gallbladder stones and bile duct dilation. The bile duct may not be dilated early in the course of acute biliary

obstruction, however. Medical management usually improves the patient's clinical status, providing a window of approximately 24 h during which biliary drainage should be established, typically byERCP. Undue delay can result in recrudesence of overt sepsis and increased morbidity. If, in addition to Charcot's triad, shock and confusion are present (Reynolds's pentad), urgent attempts to restore biliary drainage are usually indicated.

Gallstone Pancreatitis Gallstones may cause acute pancreatitis as they pass through the ampulla of Vater, where they obstruct the pancreatic duct (and sometimes cause reflux of bile into the pancreas). The occurrence of gallstone pancreatitis usually implies passage of a stone into the duodenum, and only about 20% of patients harbor a persistent stone in the ampulla or the common bile duct. Retained stones are more common in the subset of patients with jaundice, severe pancreatitis, or superimposed ascending cholangitis.

UrgentERCPdecreases the morbidity of gallstone pancreatitis in some subsets of patients, but it remains unclear whether the benefit of ERCP is mainly attributable to treatment and prevention of ascending cholangitis or to relief of pancreatic duct obstruction. ERCP is warranted early in the course of gallstone pancreatitis if ascending cholangitis is also suspected, especially in a jaundiced patient. Urgent ERCP may also be indicated in the minority of patients predicted to have severe pancreatitis using a multifactorial index of severity such as the Glasgow, Ranson's, or Apache II score.

ELECTIVE ENDOSCOPY

Dyspepsia and Reflux Dyspepsia is a burning discomfort in the upper abdomen that may be caused by diverse processes such as gastroesophageal reflux, peptic ulcer disease, and "nonulcer dyspepsia," a heterogeneous category that includes disorders of motility, sensation, and somatization. Gastric and esophageal malignancies are less common causes of dyspepsia. Careful history taking allows accurate differential diagnosis of dyspepsia in only about half of patients. In the remainder, endoscopy can be a useful diagnostic tool, especially in those patients whose symptoms are not resolved by an empirical trial of symptomatic treatment.

Gastroesophageal Reflux Disease (GERD) When classic symptoms of gastroesophageal reflux are present, such as water brash and substernal heartburn, presumptive diagnosis and empirical treatment are often sufficient. Although endoscopy is sensitive for diagnosis of esophagitis, it misses some cases of reflux, since some patients have symptomatic reflux without esophagitis. The most sensitive test for diagnosis of GERD is 24-h ambulatory pH monitoring. Endoscopy is nevertheless indicated in patients with resistant reflux symptoms and in those with recurrent dyspepsia after treatment that is not clearly due to reflux on clinical grounds alone, to assess the esophagus and exclude other diseases. Endoscopy is also advised in a patient with reflux and dysphagia, to look for a stricture or malignancy. Endoscopy is probably also indicated in patients with long-standing (³10 years) frequent heartburn, who are at sixfold increased risk of Barrett's esophagus compared to a patient with<1 year of reflux symptoms. Patients with Barrett's esophagus usually enter a program of periodic endoscopy with biopsies, to detect dysplasia or early carcinoma.

Peptic Ulcer Peptic ulcer classically causes epigastric gnawing or burning, often occurring nocturnally and promptly relieved by food or antacids. Although endoscopy is the most sensitive diagnostic test for peptic ulcer, immediate endoscopy is not a cost-effective strategy in young patients with ulcer-like dyspeptic symptoms unless endoscopy is available at low cost. Patients with suspected peptic ulcer should be evaluated for *Helicobacter pylori* infection. Serology (which documents past or present infection) and urea breath testing (which demonstrates current infection) are less invasive and costly than endoscopy with biopsy. Patients with ulcer-like symptoms despite treatment should undergo endoscopy to exclude gastric malignancy, and patients with "alarm symptoms" (early satiety or anorexia, early recurrence of symptoms, anemia) should also undergo endoscopy.

Nonulcer Dyspepsia This may be associated with bloating and, unlike peptic ulcer, tends not to remit and recur. Most patients do not respond to acid-reducing, prokinetic, or anti-*Helicobacter* therapy and are referred for endoscopy to exclude a refractory ulcer. While endoscopy usefully excludes other diagnoses, it generally does little to improve the treatment of patients with nonulcer dyspepsia.

Dysphagia About 50% of patients with difficulty swallowing have a mechanical obstruction; the remainder have a motor disorder. Careful history taking often suggests a diagnosis and leads to the appropriate use of diagnostic tests. Esophageal strictures typically cause progressive dysphagia, first for solids, then liquids; esophageal motor disorders often cause intermittent dysphagia for both solids and liquids. Some underlying disorders have characteristic historical features: Schatzki's ring causes episodic dysphagia for solids, typically at the beginning of a meal; pharyngeal motor disorders are associated with difficulty initiating deglutition ("transfer dysphagia") and nasal reflux with swallowing; and achalasia may cause nocturnal regurgitation of undigested food particles.

When mechanical obstruction is suspected, endoscopy is a useful initial diagnostic test, since it permits immediate biopsy and dilation of strictures, masses, or rings. Blind or forceful passage of an endoscope may lead to perforation in a patient with stenosis of the cervical esophagus or a Zencker's diverticulum, but gentle passage of an endoscope under direct visual guidance is reasonably safe even in these patients. Endoscopy can miss a subtle stricture or ring in some patients.

When a motor disorder is suspected, esophageal radiography is the best initial diagnostic test. The pharyngeal swallowing mechanism, esophageal peristalsis, and the lower esophageal sphincter can all be assessed. In some disorders, subsequent esophageal manometry may also be important for diagnosis.

Anemia and Occult Blood in the Stool Iron-deficiency anemia may be attributed to poor iron absorption (as in celiac sprue) or, more commonly, chronic blood loss. Intestinal bleeding should be strongly suspected in men and postmenopausal women with iron-deficiency anemia, and colonoscopy is indicated in such patients, even in the absence of detectable occult blood in the stool. About 30% will have large colonic polyps, 10% will have colorectal cancer, and additional patients will have colonic vascular lesions. When a convincing source of blood loss is not found in the colon, upper gastrointestinal endoscopy should also be performed; if no lesion is found,

duodenal biopsies should be obtained to exclude sprue. Evaluation of the small bowel may be appropriate if bothEGD and colonoscopy are unrevealing.

Tests for occult blood in the stool detect hemoglobin or the heme moiety and are most sensitive for colonic blood loss, although they will also detect larger amounts of upper gastrointestinal bleeding. Patients with occult blood in normal-appearing stool should undergo colonoscopy to diagnose or exclude colorectal neoplasia. The diagnostic yield is lower than in iron-deficiency anemia. Whether upper endoscopy is also indicated largely depends on the patient's symptoms.

The small intestine may be the source of chronic intestinal bleeding, especially if colonoscopy and upper endoscopy are not diagnostic. The utility of small-bowel evaluation varies with the clinical setting and is most important in patients whose bleeding causes chronic or recurrent anemia. While small-bowel radiography is usually normal, partial or total small-bowel enteroscopy yields a specific diagnosis in about 50% of such patients. The commonest finding is mucosal vascular ectasias or telangiectasias.

Colorectal Cancer Screening Most colon cancers develop from preexisting colonic adenomas, and colorectal cancer can be largely prevented by the detection and removal of colonic adenomatous polyps. Screening for polyps and early, asymptomatic cancers can be accomplished both by testing stool specimens for occult blood and by directly examining the colonic mucosa. Since tests for occult blood are insensitive, detecting only about one-fourth of colon cancers and large polyps, visualization of at least a part of the colon is an important component of colorectal cancer screening.

The choice of screening strategy for an asymptomatic patient depends in part on their personal and family history. A past history of inflammatory bowel disease or colorectal polyps, a family history of two or more first-degree family members with adenomatous polyps or cancer, certain familial cancer syndromes, or the finding of occult blood in the stool all place an individual at increased risk and alter screening recommendations. An individual without these factors is generally considered at average risk, and screening flexible sigmoidoscopy every 5 years beginning at age 50 is recommended. Screening strategies for higher risk patients are in Table 283-2. Screening strategies for the patient with one family member with colorectal cancer are debated. When the index case occurred at a young age (<60 years), screening colonoscopy should be offered when the patient is 10 years younger than the affected relative was when diagnosed.

Flexible sigmoidoscopy is an effective screening tool for two reasons: (1) the majority of colorectal cancers have traditionally occurred in the rectum and left colon, and (2) many right-sided colon cancers are associated with synchronous left-sided adenomas. The detection of an adenoma during sigmoidoscopy generally leads to full colonoscopy and detection of right-sided cancers, if present. Over the past several decades, however, there has been a gradual change in the distribution of colon cancers, with proportionally fewer rectal and left-sided cancers than in the past. This has spurred interest in evaluating the entire colon during a single screening examination. Barium enema has been advocated but requires flexible sigmoidoscopy also, to exclude missed rectal lesions. Large studies of colonoscopy for screening of average-risk individuals are currently underway. In addition, the new imaging technique of "virtual colonoscopy"

holds considerable promise. This modality uses data from helicalCT to generate a graphical display of a "flight" down the colonic lumen. While this technique is not yet sufficiently sensitive for routine clinical use, further refinement may result in a useful noninvasive screening method.

Diarrhea Most cases of diarrhea are acute, self-limited, and due to infections or medication. Chronic diarrhea (lasting >6 weeks) is more often due to a primary inflammatory or malabsorptive disorder, is less likely to resolve spontaneously, and generally requires diagnostic evaluation. Patients with chronic diarrhea or severe, unexplained acute diarrhea often undergo endoscopy if stool tests for pathogens are unrevealing. The choice of endoscopic test depends on the clinical setting.

Patients with colonic symptoms and findings such as bloody diarrhea, tenesmus, fever, or leukocytes in stool generally undergo sigmoidoscopy or colonoscopy to look for colitis. Sigmoidoscopy is often adequate and is the best initial test in most such patients. On the other hand, patients with symptoms and findings suggesting small-bowel disease such as large-volume watery stools; substantial weight loss; and malabsorption of iron, calcium, or fat may undergo upper endoscopy with duodenal biopsies.

Many patients with chronic diarrhea do not fit either of these patterns. When there is a long-standing history of alternating constipation and diarrhea dating to early adulthood, without findings such as blood in the stool or anemia, a diagnosis of irritable bowel syndrome may be made without direct visualization of the bowel. Steatorrhea and upper abdominal pain may prompt evaluation of the pancreas rather than the gut. Patients whose chronic diarrhea is not easily categorized often undergo initial colonoscopy to examine the entire colon (and terminal ileum) for inflammatory or neoplastic disease.

Minor Hematochezia Bright red blood passed with or on formed brown stool usually has a rectal, anal, or distal sigmoid source. Patients with even trivial amounts of hematochezia should be investigated with flexible sigmoidoscopy to exclude large polyps or cancers in the distal bowel. Patients who report red blood on the toilet tissue only, without blood in the toilet or on the stool, are bleeding from a lesion in the anal canal, and careful external and digital examinations and anoscopy are sufficient for diagnosis in most cases.

Unexplained Pancreatitis About 20% of patients with pancreatitis have no identified cause after routine clinical investigation (including a review of medication and alcohol use, measurement of serum triglyceride and calcium levels, abdominal ultrasonography, and CT). Endoscopic techniques lead to a specific diagnosis in the majority of such patients, often altering clinical management. Endoscopic investigation is particularly appropriate if the patient has had more than one episode of pancreatitis.

Microlithiasis, or the presence of microscopic crystals in bile, is a leading cause of previously unexplained acute pancreatitis and is sometimes seen during abdominal ultrasonography as layering sludge or flecks of floating, echogenic material in the gallbladder. Gallbladder bile can be obtained for microscopic analysis by administering a cholecystokinin analogue during endoscopy, causing contraction of the gallbladder. Bile is suctioned from the duodenum as it drains from the papilla, and the darkest fraction is examined for cholesterol crystals or bilirubinate granules. Alternatively, bile

can be aspirated from the bile duct duringERCP or the gallbladder can be examined for sludge or crystals byEUSbefore administering cholecystokinin. The latter strategy is probably the most sensitive means of diagnosing microlithiasis.

Previously undetected chronic pancreatitis, pancreatic malignancy, or pancreas divisum may be diagnosed by eitherERCP orEUS. Although ERCP remains the gold standard imaging test for chronic pancreatitis, EUS has good sensitivity and less risk than ERCP. Sphincter of Oddi dysfunction probably causes some cases of pancreatitis and can be diagnosed by manometric studies performed during ERCP.

OPEN-ACCESS ENDOSCOPY

While gastroenterologists have traditionally seen patients in consultation before arranging an endoscopic procedure, direct scheduling of endoscopic procedures by primary care physicians, or *open-access endoscopy*, is an increasingly common practice. When the indications for endoscopy are clear cut and appropriate, the procedural risks are low, and the patient understands what to expect, open-access endoscopy streamlines patient care and decreases costs.

Patients referred for open-access endoscopy should have a recent history, physical examination, and medication review. A copy of such an evaluation should be available when the patient comes to the endoscopy suite. Patients with unstable cardiovascular or respiratory conditions should not be referred directly for open-access endoscopy. Patients with selected cardiac conditions undergoing certain procedures should be prescribed prophylactic antibiotics prior to endoscopy, as described inTable 283-1. In addition, patients taking anticoagulants may need changes in treatment before endoscopy, as detailed inTable 283-3. While many endoscopists recommend discontinuing aspirin for 5 days before elective endoscopic procedures, most evidence suggests that in the absence of a preexisting bleeding disorder it is safe to perform endoscopic procedures in patients taking aspirin and nonsteroidal anti-inflammatory drugs.

Common indications for open-accessEGDinclude dyspepsia resistant to a trial of appropriate therapy; dysphagia or odynophagia; gastrointestinal bleeding; and persistent vomiting, anorexia, or early satiety. Open-access colonoscopy is often requested in men or postmenopausal women with iron-deficiency anemia, patients with occult blood in the stool, patients with a previous history of colorectal adenomatous polyps or cancer, and for screening in patients with above-average risk for colon cancer, as described in Table 283-2. Flexible sigmoidoscopy is commonly performed as an open-access procedure for cancer screening in asymptomatic persons over 50 and for patients with hematochezia.

When patients are referred for open-access colonoscopy, the primary care provider may need to choose a colonic preparation. Commonly used oral preparations include polyethelene glycol lavage solution and sodium phosphate. Sodium phosphate may cause fluid and electrolyte abnormalities, especially in patients with renal failure, congestive heart failure, and patients over 70 years of age.

(Bibliography omitted in Palm version)

Back to Table of Contents

284. DISEASES OF THE ESOPHAGUS - *Raj K. Goyal*

The two major functions of the esophagus are the transport of the food bolus from the mouth to the stomach and the prevention of retrograde flow of gastrointestinal contents. The transport function is achieved by peristaltic contractions in the pharynx and esophagus associated with relaxation of upper and lower esophageal sphincters (Chap. 40). Retrograde flow is prevented by the two esophageal sphincters, which remain closed between swallows. The upper esophageal sphincter (UES) consists of the cricopharyngeus and inferior pharyngeal constrictor muscles, striated muscles innervated by excitatory somatic lower motor neurons. These muscles exhibit no myogenic tone and receive no inhibitory innervation. The UES remains closed owing to the elastic properties of its wall and to neurogenic tonic contraction of the sphincter muscles. It is opened by central inhibition of the sphincter muscles in concert with forward displacement of the larynx by the suprahyoid muscles. In contrast, the lower esophageal sphincter (LES) is composed of smooth muscle and is innervated by parallel sets of parasympathetic excitatory and inhibitory pathways. It remains closed because of its intrinsic myogenic tone, which is modulated by the excitatory and inhibitory nerves. It opens in response to the activity of the inhibitory nerves. The neurotransmitters of the excitatory nerves are acetylcholine and substance P, and those of the inhibitory nerves are vasoactive intestinal peptide (VIP) and nitric oxide. The function of the LES is supplemented by the striated muscle of the diaphragmatic crura, which surrounds the LES and acts as an external LES. Relaxation of the LES without esophageal contraction occurs during belching and gastric distention. Gastric distention-evoked transient lower esophageal sphincter relaxation (tLESR) is a vasovagal reflex. Fatty meals, smoking, and beverages with a high xanthine content (tea, coffee, cola) also cause a reduction in sphincter pressure. Many hormones and neurotransmitters can modify LES pressure. Muscarinic M_2 and M_3receptor agonists, a-adrenergic agonists, gastrin, substance P, and prostaglandin F_{2a}cause contraction. Nicotine, b-adrenergic agonists, dopamine, cholecystokinin, secretin, VIP, calcitonin gene-related peptide (CGRP), adenosine, prostaglandin E, and nitric oxide donors such as nitrates reduce sphincter pressure.

SYMPTOMS

DYSPHAGIA See Chap. 40.

ESOPHAGEAL PAIN

Heartburn, or pyrosis, is characterized by burning retrosternal discomfort that may move up and down the chest like a wave. When severe, it may radiate to the sides of the chest, the neck, and the angles of the jaw. Heartburn is a characteristic symptom of reflux esophagitis and may be associated with regurgitation or a feeling of warm fluid climbing up the throat. It is aggravated by bending forward, straining, or lying recumbent and is worse after meals. It is relieved by an upright posture, by the swallowing of saliva or water, and, more reliably, by antacids. Heartburn is produced by heightened mucosal sensitivity and can be reproduced by infusion of dilute (0.1 *N*) hydrochloric acid (Bernstein test) or neutral hyperosmolar solutions into the esophagus.

Odynophagia, or painful swallowing, is characteristic of nonreflux esophagitis,

particularly monilial and herpes esophagitis. Odynophagia may occur with peptic ulcer of the esophagus (Barrett's ulcer), carcinoma with periesophageal involvement, caustic damage of the esophagus, and esophageal perforation. Odynophagia is unusual in uncomplicated reflux esophagitis. Crampy chest pain associated with impaction of a food bolus should be distinguished from odynophagia.

Atypical chest pain other than heartburn and odynophagia occurs in reflux esophagitis or esophageal motility disorders such as diffuse esophageal spasm. Spasm may occur spontaneously or during a meal. Chest pain due to periesophageal involvement with carcinoma or peptic ulcer may be constant and agonizing. Sometimes different types of esophageal pains exist together in the same patient, and frequently patients are not able to describe the pain accurately enough to allow its classification. Coronary artery disease should always be excluded before the esophagus is considered as the cause of atypical chest pain. The most frequent esophageal cause of chest pain is reflux esophagitis. Some patients with atypical chest pain have nonspecific esophageal motor abnormalities of uncertain significance. Many of these patients have behavioral abnormalities, psychosomatic disorders, depression, anxiety, panic reactions, and other psychological disorders.

REGURGITATION

Regurgitation is the effortless appearance of gastric or esophageal contents in the mouth. In distal esophageal obstruction and stasis, as in achalasia or the presence of a large diverticulum, the regurgitated material consists of tasteless mucoid fluid or undigested food. Regurgitation of sour or bitter-tasting material occurs in severe gastroesophageal reflux and is associated with incompetence of both theUES and the LES. Regurgitation may result in laryngeal aspiration, with spells of coughing and choking that awaken the patient from sleep, and in aspiration pneumonia. Water brash is reflex salivary hypersecretion that occurs in response to peptic esophagitis and should not be confused with regurgitation.

DIAGNOSTIC TESTS

RADIOLOGIC STUDIES

Barium swallow with fluoroscopy and an esophagogram is a widely used test for the diagnosis of esophageal disease and can be used to evaluate both structural and motor disorders. Spontaneous reflux of barium from the stomach into the esophagus suggests gastroesophageal reflux. Esophageal peristalsis is best studied in the recumbent position, because in the upright position barium passage occurs largely by gravity alone. A double-contrast esophagogram, obtained by coating the esophageal mucosa with barium and distending the esophageal lumen with air using effervescent granules, is particularly useful in demonstrating mucosal ulcers and early cancers. A barium-soaked piece of bread or a 13-mm barium tablet is sometimes used to demonstrate an obstructive lesion.Figures 284-1 and284-2illustrate the radiographic appearance of some esophageal disorders. Since the oropharyngeal phase of swallowing lasts no more than a second, videofluoroscopy is necessary to permit detection and analysis of abnormalities of oral and pharyngeal function. The pharynx is examined to detect stasis of barium in the valleculae and piriform sinuses and regurgitation of barium into the

nose and tracheobronchial tree.

ESOPHAGOSCOPY

Esophagoscopy is the direct method of establishing the cause of mechanical dysphagia and of identifying mucosal lesions that may not be identified by the usual barium swallow. If the lumen is markedly narrowed, use of a smaller-caliber endoscope may be needed; on occasion a stricture must be dilated before the examination can be completed. Endoscopic biopsies are useful in diagnosing carcinoma, reflux esophagitis, and other mucosal diseases. Cells obtained by a cytology balloon or brushing the mucosa can be evaluated for carcinoma. Endoscopic ultrasonography permits evaluation of intramural masses and staging of esophageal cancer.

ESOPHAGEAL MOTILITY

The study of esophageal motility entails simultaneous recording of pressures from different sites in the esophageal lumen with an assembly of pressure sensors positioned 5 cm apart. TheUES andLESappear as zones of high pressure that relax on swallowing. The pharynx and esophagus normally show peristaltic waves with each swallow.

Esophageal motility studies are helpful in the diagnosis of esophageal motor disorders (achalasia, spasm, scleroderma) (Fig. 284-3) but are of little value in the diagnosis of mechanical dysphagia. In patients with reflux esophagitis, esophageal manometry is useful in quantitating lower esophageal competence and providing information on the status of the esophageal body motor activity. Manometry provides quantitative data that cannot be obtained by barium swallow or endoscopy. Tests for reflux esophagitis are described later.

MOTOR DISORDERS

STRIATED MUSCLE

Oropharyngeal Paralysis Paralysis of oral muscle leads to difficulty initiating swallowing and drooling of food out of the mouth. Pharyngeal paralysis, characterized by dysphagia, nasal regurgitation, and aspiration during swallowing, occurs in a variety of neuromuscular disorders (see Table 40-1). Some of these disorders also involve laryngeal muscles, causing hoarseness. When the suprahyoid muscles are paralyzed, theUES does not open with swallowing, leading to paralytic achalasia of the UES and severe dysphagia.

Videofluoroscopy with barium of various consistencies may reveal difficulties in the oral phase of swallowing. The test may show barium in the valleculae and piriform sinuses, nasal and tracheal aspiration, failure of the upper sphincter to open, and/or abnormal movement of the hyoid bone and the larynx with a swallow (Fig. 284-1). Motility studies demonstrate a reduced amplitude of pharyngeal and upper esophageal contractions and reduced basal upper esophageal sphincter pressure without further relaxation on swallowing (Fig. 284-3). Patients with myasthenia gravis (Chap. 380) and polymyositis (Chap. 382) respond to treatment. Dysphagia resulting from a cerebrovascular accident improves with time, although often not completely. Treatment consists of maneuvers to

reduce pharyngeal stasis and enhance airway protection under the direction of a trained swallow therapist. Feeding by a nasogastric tube or an endoscopically placed gastrostomy tube may be necessary for nutritional support; however, these maneuvers do not provide protection against aspiration of salivary secretions. Cricopharyngeal myotomy is sometimes performed, but its usefulness is unproven. Extensive operative procedures to prevent aspiration are rarely needed. Death is often due to pulmonary complications.

Cricopharyngeal Bar Failure of the cricopharyngeus to relax on swallowing appears as a prominent bar on the posterior wall of the pharynx on barium swallow (Fig. 284-1). A transient cricopharyngeal bar is seen in up to 5% of individuals without dysphagia undergoing upper gastrointestinal studies; it can be produced in normal individuals during a Valsalva maneuver. A persistent cricopharyngeal bar may be caused by fibrosis in the cricopharyngeus. Some of these patients complain of food sticking in their throats. Cricopharyngeal myotomy may be helpful but is contraindicated in the presence of gastroesophageal reflux because it may lead to pharyngeal and pulmonary aspiration.

Globus Pharyngeus A sensation of a constant lump in the throat, but no difficulty in swallowing, occurs especially in individuals with emotional disorders, particularly women. Results of barium studies and manometry are normal. Treatment consists primarily of reassurance. Some patients with globus pharyngeus have associated reflux esophagitis, and they may respond to treatment of the esophagitis.

SMOOTH MUSCLE

Achalasia Achalasia is a motor disorder of the esophageal smooth muscle in which theLES does not relax normally with swallowing, and the esophageal body undergoes nonperistaltic contractions.

Pathophysiology The underlying abnormality is the loss of intramural neurons. Inhibitory neurons containingVIP and nitric oxide synthase are predominantly involved, but in advanced disease cholinergic neurons are also affected. Primary idiopathic achalasia accounts for most of the patients seen in the United States. Secondary achalasia may be caused by gastric carcinoma that infiltrates the esophagus, lymphoma, Chagas' disease, certain viral infections, eosinophilic gastroenteritis, and neurodegenerative disorders.

Clinical features Achalasia affects patients of all ages and both sexes. Dysphagia, chest pain, and regurgitation are the main symptoms. Dysphagia appears early, occurs with both liquids and solids, and is worsened by emotional stress and hurried eating. Various maneuvers designed to increase intraesophageal pressure, including the Valsalva maneuver, may aid the passage of the bolus into the stomach. Regurgitation and pulmonary aspiration occur because of retention of large volumes of saliva and ingested food in the esophagus. Patients may complain of difficulty belching. The presence of gastroesophageal reflux argues against achalasia; and in patients with long-standing heartburn, cessation of heartburn and appearance of dysphagia suggest development of achalasia on top of reflux esophagitis. The course is usually chronic, with progressive dysphagia and weight loss over months to years. Achalasia associated with carcinoma

is characterized by severe weight loss and a rapid downhill course if untreated.

Diagnosis A chest x-ray shows absence of the gastric air bubble and sometimes a tubular mediastinal mass beside the aorta. An air-fluid level in the mediastinum in the upright position represents retained food in the esophagus. Barium swallow shows esophageal dilation, and in advanced cases the esophagus may become sigmoid. On fluoroscopy, normal peristalsis is lost in the lower two-thirds of the esophagus. The terminal part of the esophagus shows a persistent beaklike narrowing representing the nonrelaxingLES(Fig. 284-1).

Manometry shows the basalLESpressure to be normal or elevated, and swallow-induced relaxation either does not occur or is reduced in degree, duration, and consistency. The esophageal body shows an elevated resting pressure. In response to swallows, primary peristaltic waves are replaced by simultaneous-onset contractions (Fig. 284-3). These contractions may be of poor amplitude (classic achalasia) or of large amplitude and long duration (vigorous achalasia). Cholecystokinin (CCK), which normally causes a fall in the sphincter pressure, paradoxically causes contraction of the LES (the CCK test). This paradoxical response occurs because, in achalasia, the neurally transmitted inhibitory effect of CCK is absent owing to the loss of inhibitory neurons. Endoscopy is helpful in excluding the secondary causes of achalasia, particularly gastric carcinoma.

TREATMENT

Treatment with soft foods, sedatives, and anticholinergic drugs is usually unsatisfactory. Nitrates and calcium channel blockers provide short-term benefit, but their use may be limited by side effects. Nitroglycerin, 0.3 to 0.6 mg, is used sublingually before meals and as needed for chest pain. Isosorbide dinitrate, 2.5 to 5 mg sublingually or 10 to 20 mg orally, is used before meals. Nitrates are associated with headache and postural hypotension. The calcium channel blocker nifedipine, 10 to 20 mg orally or sublingually before meals, is also effective. Endoscopic intrasphincteric injection of botulinum toxin is effective over a short period in some patients. Repeated injections may lead to fibrosis, complicating further operative therapy. Botulinum toxin acts by blocking cholinergic excitatory nerves in the sphincter. Balloon dilatation reduces the basalLESpressure by tearing muscle fibers. In experienced hands, this technique is effective in ~85% of patients. Perforation and bleeding are potential complications. Heller's extramucosal myotomy of the LES, in which the circular muscle layer is incised, is equally effective. Laparoscopic myotomy is the procedure of choice. Reflux esophagitis and peptic stricture may follow successful treatment (more often with myotomy than with balloon dilatation).

Diffuse Esophageal Spasm and Related Motor Disorders These disorders present with clinical symptoms of chest pain and dysphagia and are recognized by their manometric features. In pure form, they all show normal relaxation to swallows. Diffuse esophageal spasm is characterized by nonperistaltic contractions, usually of large amplitude and long duration. An esophageal motility pattern showing hypertensive but peristaltic contractions has been called "nutcracker esophagus."

Pathophysiology Nonperistaltic contractions are due to dysfunction of inhibitory nerves.

Histopathologic studies show patchy neural degeneration localized to nerve processes, rather than the prominent degeneration of nerve cell bodies seen in achalasia. Diffuse esophageal spasm may progress to achlasia. Hypertensive peristaltic contractions and hypertensive or hypercontractingLES may represent cholinergic or myogenic hyperactivity.

Clinical features Diffuse spasm and related motor disorders cannot be distinguished clinically. They all present with chest pain, dysphagia, or both. Chest pain is particularly marked in patients with esophageal contractions of large amplitude and long duration. Chest pain usually occurs at rest but may be brought on by swallowing or by emotional stress. The pain is retrosternal; it may radiate to the back, the sides of the chest, both arms, or the sides of the jaw and may last from a few seconds to several minutes. It may be acute and severe, mimicking the pain of myocardial ischemia. Dysphagia for solids and liquids may occur with or without chest pain and is correlated particularly with simultaneous-onset contractions.

Diffuse esophageal spasm and related esophageal motor disorders must be differentiated from other causes of chest pain, particularly ischemic heart disease with atypical angina. A complete cardiac workup should be done before a noncardiac etiology is considered seriously. The presence of dysphagia in association with pain should point to the esophagus as the site of disease. Esophageal motility disorders are an uncommon cause of noncardiac chest pain, which is more commonly due to reflux esophagitis or visceral hypersensitivity.

Diagnosis In diffuse esophageal spasm, barium swallow shows that normal sequential peristalsis below the aortic arch is replaced by uncoordinated simultaneous contractions that produce the appearance of curling or multiple ripples in the wall, sacculations, and pseudodiverticula -- the "corkscrew" esophagus (Fig. 284-1). Sometimes an esophageal contraction obliterates the lumen, and barium is pushed away in both directions. The barium swallow is frequently normal in diffuse esophageal spasm and mostly normal in the related disorders.

Diffuse esophageal spasm (Fig. 284-3) and related motor disorders (hypertensive peristaltic contraction, hypertensiveLES and hypercontracting LES) are manometric diagnoses. Because these abnormalities may be episodic, the results of manometry may be normal at the time of the study. Several techniques are used to provoke esophageal spasm. Cold swallows produce the chest pain but do not produce spasm on manometric studies. Solid boluses and pharmacologic agents, particularly edrophonium, induce both chest pain and motor abnormalities. However, correlation between induction of pain and motility changes is poor. The usefulness of pharmacologic provocative tests is limited.

TREATMENT

Anticholinergics are usually of limited value. Agents that relax smooth muscle, such as sublingual nitroglycerin (0.3 to 0.6 mg) or longer-acting agents such as isosorbide dinitrate (10 to 30 mg orally before meals) and nifedipine (10 to 20 mg orally before meals) are helpful. Sublingual forms of these agents can also be used. Reassurance and tranquilizers are helpful in allaying apprehension.

Scleroderma Esophagus The esophageal lesions in systemic sclerosis consist of atrophy of smooth muscle, manifested by weakness in the lower two-thirds of the esophageal body and incompetence of theLES. The esophageal wall is thin and atrophic and may exhibit areas of patchy fibrosis. Patients usually present with dysphagia to solids. Liquids may cause dysphagia when the patient is recumbent. These patients usually also complain of heartburn, regurgitation, and other symptoms of gastroesophageal reflux disease (GERD). Barium swallow shows dilation and loss of peristaltic contractions in the middle and distal portions of the esophagus. The LES is patulous, and gastroesophageal reflux may occur freely (Fig. 284-1). Mucosal changes due to esophageal ulceration and esophageal stricture may be present. Motility studies show a marked reduction in the amplitude of smooth-muscle contractions, which may be peristaltic or nonperistaltic. The resting pressure of the LES is subnormal, but sphincter relaxation is normal (Fig. 284-3). Similar esophageal motor abnormalities are found in other collagen vascular diseases and in Raynaud's syndrome alone. Dietary adjustments with the use of soft foods are helpful in management. GERD and its complications should be treated aggressively.

GASTROESOPHAGEAL REFLUX DISEASE

GERDis one of the most prevalent gastrointestinal disorders. Population-based studies show that up to 15% of individuals have heartburn at least once a week and about 7% have heartburn daily. Symptoms are caused by back flow of gastric acid and other gastric contents into the esophagus due to incompetent barriers at the gastroesophageal junction.

Pathophysiology The normal antireflux mechanisms consist of theLES, the crural diaphragm, and the anatomic location of the gastroesophageal junction below the diaphragmatic hiatus. Reflux occurs only when the gradient of pressure between the LES and the stomach is lost. It can be caused by a sustained or transient decrease in LES tone. A sustained hypotension of the LES may be due to muscle weakness that is often without apparent cause. Secondary causes of LES incompetence include scleroderma-like diseases, myopathy associated with chronic intestinal pseudo-obstruction, pregnancy, smoking, anticholinergic drugs, smooth-muscle relaxants [b-adrenergic agents, aminophylline, nitrates, calcium channel blockers, phosphodiesterase inhibitors that increase cyclic AMP or cyclic GMP (including sildenofil)], surgical destruction of the LES, and esophagitis.tLESRwithout associated esophageal contraction is due to a vagal reflex in which LES relaxation is elicited by gastric distention. Increased tLESR is associated withGERD. A similar reflex operates during belching. Apart from incompetent barriers, gastric contents are most likely to reflux (1) when gastric volume is increased (after meals, in pyloric obstruction, in gastric stasis, during acid hypersecretion states), (2) when gastric contents are near the gastroesophageal junction (in recumbency, bending down, hiatus hernia), and (3) when gastric pressure is increased (obesity, pregnancy, ascites, tight clothes). Incompetence of the diaphragmatic crural muscle, which surrounds the esophageal hiatus in the diaphragm and functions as an external LES, also predisposes to GERD.

The total exposure of the esophagus to refluxed acid correlates with potential for mucosal damage. Exposure depends on the amount of refluxed material per episode,

frequency of episodes, and rate of clearing the esophagus by gravity and peristaltic contractions. When peristaltic contractions are impaired, esophageal clearance is impaired. Acid refluxed into the esophagus is neutralized by saliva. Thus, impaired salivary secretion also increases esophageal exposure time. If the refluxed material extends to the cervical esophagus and breaches the upper sphincter, it can enter the pharynx, larynx, and trachea, causing chronic cough, bronchoconstriction, pharyngitis, laryngitis, or bronchitis.

Reflux esophagitis is a complication of reflux and develops when mucosal defenses are unable to counteract the damage done by acid, pepsin, and bile. *Mild esophagitis* involves microscopic changes of mucosal infiltration with granulocytes or eosinophils, hyperplasia of basal cells, and elongation of dermal pegs. Endoscopic appearance may be normal. *Erosive esophagitis* involves endoscopically apparent mucosal damage, redness, friability, bleeding, superficial, linear ulcers, and exudates. *Peptic stricture* results from fibrosis that causes lumenal constriction. These strictures occur in ~10% of patients with untreatedGERD. Short strictures caused by spontaneous reflux are usually 1 to 3 cm long and are present in the distal esophagus near the squamocolumnar junction (Fig. 284-2). Long, tubular peptic strictures can result from persistent vomiting or prolonged nasogastric intubation. Erosive esophagitis may cause bleeding and heal by intestinal metaplasia (*Barrett's esophagus*) that is a risk factor for adenocarcinoma.

Clinical Features Regurgitation of sour material in the mouth and heartburn are the characteristic symptoms ofGERD. Heartburn is produced by the contact of refluxed material with the inflamed or sensitized esophageal mucosa. Angina-like or atypical chest pain occurs in some patients, while others experience no heartburn or chest pain. Persistent dysphagia suggests development of a peptic stricture. Most patients with peptic stricture have a history of several years of heartburn preceding dysphagia. However, in one-third of patients, dysphagia is the presenting symptom. Rapidly progressive dysphagia and weight loss may indicate the development of adenocarcinoma in Barrett's esophagus. Bleeding occurs due to mucosal erosions or Barrett's ulcer. Severe reflux may reach the pharynx and mouth and result in laryngitis, morning hoarseness, and pulmonary aspiration. Recurrent pulmonary aspiration can cause aspiration pneumonia, pulmonary fibrosis, or chronic asthma. By contrast, many patients with GERD remain asymptomatic or self-treated and do not seek attention until severe complications occur.

Diagnosis The diagnostic approach toGERD can be divided into three categories:

1. documentation of mucosal injury,

2. documentation and quantitation of reflux, and

3. definition of the pathophysiology.

Reflux esophagitis and its complications are documented by the use of barium swallow, esophagoscopy, and mucosal biopsy. The results of barium swallow are usually normal in uncomplicated esophagitis but may reveal a stricture or ulcer. A high esophageal peptic stricture, a deep ulcer, or adenocarcinoma suggest Barrett's esophagus. Uncomplicated Barrett's esophagus is not diagnosed reliably by barium studies.

Esophagoscopy may reveal the presence of erosive esophagitis, distal peptic stricture, or a columnar-cell-lined lower esophagus with or without a proximally located peptic stricture, ulcer, or adenocarcinoma. Results of esophagoscopy may be normal in many patients with esophagitis; in such patients, mucosal biopsies and the Bernstein test are helpful. The mucosal biopsies should be performed at least 5 cm above theLES, because the esophageal mucosal changes of chronic esophagitis are quite frequent in the most distal esophagus in otherwise normal individuals. About 10% of biopsies yield a false-positive or false-negative result. The Bernstein test involves the infusion of solutions of 0.1 *N* HCl and normal saline into the esophagus. It is useful in diagnosing reflux esophagitis that is not endoscopically obvious. In patients with reflux esophagitis, infusion of acid, but not of saline, reproduces the symptoms of heartburn. Infusion of acid in normal individuals usually produces no symptoms. Supraesophageal manifestations are diagnosed by careful otolaryngological exam.

A therapeutic trial with a proton pump inhibitor (such as omeprazole, 40 mg bid) for 1 week provides strong support for the diagnosis ofGERD.

Documentation and quantitation of reflux when necessary can be done by ambulatory long-term (24-h) esophageal pH recording. For evaluation of pharyngeal reflux, a system of recording simultaneously from pharyngeal and esophageal sites may be useful. The pH recordings are helpful only in the evaluation of acid reflux. The presence of bile or intestinal alkaline secretions is suggested by the occurrence of reflux symptoms in the absence of gastric acid and demonstration of bile in an aspirate of esophageal reflux fluid. Documentation of reflux is necessary only when the role of reflux in the symptom complex is unclear, particularly in evaluation of supraesophageal symptoms and chest pain without endoscopic evidence of esophagitis.

Definition of pathophysiologic factors inGERD is sometimes indicated for management decisions such as antireflex surgery. Esophageal motility studies may provide useful quantitative information on the competence of theLES and on esophageal motor function.

TREATMENT

The goals of treatment are to decrease gastroesophageal reflux, render the refluxate harmless, improve esophageal clearance, and protect the esophageal mucosa. The management of uncomplicated cases generally includes weight reduction, sleeping with the head of the bed elevated by about 4 to 6 in. with blocks, and elimination of factors that increase abdominal pressure. Patients should not smoke and should avoid consuming fatty foods, coffee, chocolate, alcohol, mint, orange juice, and certain medications (such as anticholinergic drugs, calcium channel blockers, and other smooth-muscle relaxants). They should also avoid ingesting large quantities of fluids with meals. In mild cases, life-style changes and over-the-counter antisecretory agents may be adequate. In moderate cases, H_2receptor blocking agents (cimetidine, 300 mg; ranitidine, 150 mg bid; famotidine, 20 mg bid; nizatidine 150 mg bid) for 6 to 12 weeks are effective in symptom relief. Higher doses are necessary for healing erosive esophagitis, but proton pump inhibitors (PPIs) are more effective in this setting.

In cases resistant to H_2receptor blockers and severe cases, rigorous acid suppression

with aPPI is recommended. The PPIs are comparably effective: omeprazole (40 mg/d), lansoprazole (30 mg/d), pantoprazole (40 mg/d), and rabeprazole (20 mg/d) for 8 weeks can heal erosive esophagitis in up to 90% of patients. Reflux esophagitis requires prolonged therapy, for 3 to 6 months or longer if the disease recurs quickly. After initial therapy, a lower maintenance dose of PPI is used. Side effects are minimal. Aggressive acid suppression causes hypergastrinemia but does not increase the risk for carcinoid tumors or gastrinomas. Vitamin B_{12}absorption is compromised by the treatment. Patients with reflux esophagitis who have complications, such as Barrett's esophagus with concomitant esophagitis, should be treated vigorously. Patients who have an associated peptic stricture are treated with dilators to relieve dysphagia as well as provided with vigorous treatment for reflux.

Antireflux surgery, in which the gastric fundus is wrapped around the esophagus (fundoplication), increases theLESpressure and should be considered for patients with resistant and complicated reflux esophagitis that does not respond fully to medical therapy or for patients for whom long-term medical therapy is not desirable. Laparoscopic fundoplication is the surgery of choice. Ideal candidates for fundoplication are those in whom motility studies show persistently inadequate LES pressure but normal peristaltic contractions in the esophageal body.

Patients with alkaline esophagitis are treated with general antireflux measures and neutralization of bile salts with cholestyramine, aluminum hydroxide, or sucralfate. Sucralfate is particularly useful in these cases, as it also serves as a mucosal protector.

BARRETT'S ESOPHAGUS

The metaplasia of esophageal squamous epithelium to columnar epithelium (Barrett's esophagus) is a complication of severe reflux esophagitis, and it is a risk factor for esophageal adenocarcinoma (Chap. 90). Metaplastic columnar epithelium develops during healing of erosive esophagitis with continued acid reflux because columnar epithelium is more resistant to acid-pepsin damage than squamous epithelium. The metaplastic epithelium is a mosaic of different epithelial types including goblet cells and columnar cells that have features of both secretory and absorptive cells (incomplete or type III metaplasia). Barrett's epithelium progresses through a dysplastic stage before developing into adenocarcinoma. The rate of cancer development is 1 in 200 patient years; those with longer than 2 to 3 cm of intestinal metaplasia have a risk of developing esophageal cancer that is 30 to 125 times the risk of the general population.

Given the natural history, reflux esophagitis should be aggressively treated with drugs, and erosive esophagitis should be treated with drugs and surgery, if necessary, to prevent Barrett's esophagus. The prevalence of intestinal metaplasia is estimated at 4 to 10% of patients with significant heartburn. Barrett's esophagus is more common in men, particularly white men, and prevalence increases with age. A one-time esophagoscopy is recommended in patients with persistentGERDsymptoms at age 50. Established metaplasia does not regress with treatment; thus, acid suppression and fundoplication are indicated only when active esophagitis is also present.

The need and frequency of surveillance endoscopies in patients with established Barrett's esophagus are debated. The risk of developing esophageal adenocarcinoma is

related to the length of involved esophageal mucosa. People with short segments of Barrett's esophagus (distal 2 to 3 cm) account for up to 25% of unselected patients undergoing endoscopy with or withoutGERDsymptoms and appear to be at low risk. They are not routinely surveyed. However, those with long-segment Barrett's esophagus (>3 cm) are advised to have endoscopic surveillance at 1-year intervals for 2 years and then every 2 to 3 years. The frequency is increased if dysplasia is detected independent of the length of the metaplasia. Optical methods of recognizing dysplasia during the endoscopy (laser-induced fluorescence spectroscopy, optical coherence tomography) are being developed. Once high-grade dysplasia is detected, treatment of choice is esophagectomy of the Barrett's segment. Photodynamic laser or thermocoagulative mucosal ablation and endoscopic mucosal resection are being evaluated as alternatives.

Barrett's esophagus can also lead to chronic peptic ulcer of the esophagus with high (midesophageal) and long strictures.

INFLAMMATORY DISORDERS

INFECTIOUS ESOPHAGITIS

Infectious esophagitis can be due to viral, bacterial, fungal, or parasitic organisms. In severely immunocompromised patients, multiple organisms may coexist.

Viral Esophagitis *Herpes simplex virus* (HSV) type 1 occasionally causes esophagitis in immunocompetent individuals, but either HSV type 1 or HSV type 2 may afflict patients who are immunosuppressed (Chap. 182). Patients complain of an acute onset of chest pain, odynophagia, and dysphagia. Bleeding may occur in severe cases; and systemic manifestations such as nausea, vomiting, fever, chills, and mild leukocytosis may be present. Herpetic vesicles on the nose and lips may provide a clue to the diagnosis. Barium swallow is inadequate to detect early lesions and cannot reliably distinguish HSV infection from other types of infections. Endoscopy shows vesicles and small, discrete, punched-out superficial ulcerations with or without a fibrinous exudate. In later stages, a diffuse erosive esophagitis develops from enlargement and coalescence of the ulcers. Mucosal cells from a biopsy sample taken at the edge of an ulcer or from a cytologic smear show ballooning degeneration, ground-glass changes in the nuclei with eosinophilic intranuclear inclusions (Cowdry type A), and giant cell formation on routine stains. Culture for HSV becomes positive within days and is helpful in diagnosis. In patients with severe odynophagia, intravenous acyclovir, 400 mg five times a day, is usually initiated. Symptoms usually resolve in 1 week, but large ulcerations may take longer to heal. Foscarnet (90 mg/kg intravenously every 8 h) is used if resistance to acyclovir occurs.

Varicella-zoster virus (VZV) (Chap. 183) sometimes produces esophagitis in children with chickenpox and adults with herpes zoster. Esophageal VZV also can be the source of disseminated VZV infection without skin involvement. In an immunocompromised host, VZV esophagitis causes vesicles and confluent ulcers and usually resolves spontaneously, but it may cause necrotizing esophagitis in a severely compromised host. On routine histologic examination of mucosal biopsy samples or cytology specimens, VZV is difficult to distinguish fromHSV, but the distinction can be made

immunohistologically or by culture. Acyclovir reduces the duration of symptoms in VZV esophagitis.

Cytomegalovirus (CMV) infections (Chap. 185) occur only in immunocompromised patients. CMV is usually activated from a latent stage or may be acquired from blood product transfusions. CMV lesions initially appear as serpiginous ulcers in an otherwise normal mucosa. These may coalesce to form giant ulcers, particularly in the distal esophagus.

Patients present with odynophagia, chest pain, hematemesis, nausea, and vomiting. Diagnosis requires endoscopy and biopsies of the ulcer. Mucosal brushings are not useful. Routine histologic examination shows intranuclear and small intracytoplasmic inclusions in large fibroblasts and endothelial cells. Immunohistology with monoclonal antibodies toCMV and in situ hybridization of CMV DNA on centrifugation culture and are useful for early diagnosis. Ganciclovir, 5 mg/kg every 12 h intravenously, is the treatment of choice. Foscarnet (90 mg/kg every 12 h intravenously) is used in resistant cases. Therapy is continued until healing occurs, which may take 2 to 4 weeks.

HIV (Chap. 309) may be associated with a self-limited syndrome of acute esophageal ulceration associated with oral ulcers and a maculopapular skin rash, which occurs at the time of HIV seroconversion. Some patients with advanced disease have deep, persistent esophageal ulcers requiring treatment with oral glucocorticoids or thalidomide. Some ulcers respond to local steroid injection.

Bacterial Esophagitis *Bacterial esophagitis* is unusual, but esophagitis caused by *Lactobacillus* and b-hemolytic streptococci can occur in the immunocompromised host. In patients with profound granulocytopenia and patients with cancer, bacterial esophagitis is often missed because it is commonly present with other organisms, including viruses and fungi. In patients with AIDS, infection with *Cryptosporidium* or *Pneumocystis carinii* may cause nonspecific inflammation, and *Mycobacterium tuberculosis* infection may cause deep ulcerations of the distal esophagus.

***Candida* Esophagitis** *Candida* species are normal commensals in the throat but become pathogenic and produce esophagitis in immunodeficiency states. *Candida* esophagitis can occur without any predisposing factors. Patients may be asymptomatic or complain of odynophagia and dysphagia. Oral thrush or other evidence of mucocutaneous candidiasis may be absent. Rarely, *Candida* esophagitis is complicated by esophageal bleeding, perforation, and stricture or by systemic invasion. Barium swallow may be normal or show multiple nodular filling defects of various sizes (Fig. 284-2). Large nodular defects may resemble grape clusters. Endoscopy shows small, yellow-white raised plaques with surrounding erythema in mild disease. Confluent linear and nodular plaques reflect extensive disease. Diagnosis is made by demonstration of yeast or hyphal forms in plaque smears and exudate stained with periodic acid-Schiff or Gomori silver stains. Histologic examination is often negative. Culture is not useful in diagnosis but may define the species and the drug sensitivities of the yeast (Chap. 205). Fluconazole (200 mg on the first day, followed by 100 mg daily) is the preferred treatment of esophageal candidiasis because it is effective and its absorption is not affected by high gastric pH. Fluconazole is available in oral and intravenous formulations. Ketoconazole (200 to 400 mg in a single daily oral dose) is also effective

treatment, and the higher dose is used in severely immunocompromised hosts; however, its bioavailability is severely reduced at increased gastric pH. Patients who respond poorly are treated with amphotericin, 10 to 15 mg as an intravenous infusion for 6 h daily to a total dose of 300 to 500 mg. Nystatin oral suspension (100,000 units per ml) in doses of 10 to 20 mL every 6 h is effective for oral thrush. In resistant cases, amphotericin lozenges are used for 7 to 10 days followed by nystatin or fluconazole for as long as the host resistance remains low.

OTHER TYPES OF ESOPHAGITIS

Radiation esophagitis is a common occurrence during radiation treatment for thoracic cancers. The frequency and severity of esophagitis increase with the amount of radiation delivered and may be enhanced by radiosensitizing drugs like doxorubicin, bleomycin, cyclophosphamide, and cisplatin. Dysphagia and odynophagia may last several weeks to several months after therapy. The esophageal mucosa becomes erythematous, edematous, and friable. Superficial erosions coalesce to form larger superficial ulcers. Submucosal fibrosis and degenerative changes in the blood vessels, muscles, and myenteric neurons may occur. The treatment is relief of pain with viscous lidocaine during the acute phase; indomethacin treatment may reduce radiation damage. Esophageal stricture may develop.

Corrosive esophagitis is caused by the ingestion of caustic agents, such as strong alkali or acid. Severe corrosive injury may lead to esophageal perforation, bleeding, and death. Glucocorticoids are not useful in acute corrosive esophagitis. Healing is usually associated with stricture formation. Caustic strictures are usually long and rigid (Fig. 284-2) and generally require dilatation with dilators passed over a guidewire through the stricture. *Pill-induced esophagitis* is associated with the ingestion of certain types of pills and occurs most often in bedridden patients. Antibiotics such as doxycycline, tetracycline, oxytetracycline, minocycline, penicillin, and clindamycin account for more than half the cases. Nonsteroidal anti-inflammatory agents such as aspirin, indomethacin, and ibuprofen may cause injury. Other commonly prescribed pills that cause esophageal injury include potassium chloride, ferrous sulfate or succinate, quinidine, alprenolol, theophylline, ascorbic acid, pinaverum bromide, alendronate, and pamidronate. Pill esophagitis can be prevented by avoiding the offending agents or by having patients take pills in the upright position and wash them down with copious amounts of fluids.

Sclerotherapy for bleeding esophageal varices usually produces transient retrosternal chest pain and dysphagia; esophageal ulcer, stricture, hematoma, or perforation may occur. Variceal banding causes similar complications but less frequently. *Esophagitis associated with mucocutaneous and systemic diseases* is usually associated with blister and bulla formation, epithelial desquamation, and thin, weblike, or dense esophageal strictures. Pemphigus vulgaris and bullous pemphigoid form intraepithelial and subepithelial bullae, respectively, and can be distinguished by specific immunohistology; both are characterized by sloughing of epithelium or the presence of esophageal casts. Glucocorticoid treatment is usually effective. Cicatricial pemphigoid, Stevens-Johnson syndrome, and toxic epidermolysis bullosa can produce esophageal bullous lesions and strictures requiring gentle dilatation. Graft-versus-host disease occurs in patients who have received allogeneic bone marrow transplants and is associated with generalized

desquamation and esophageal strictures. Behcet's disease and eosinophilic gastroenteritis may involve the esophagus and may respond to glucocorticoid therapy. An erosive lichen planus also can involve the esophagus. Crohn's disease may cause inflammatory strictures, sinus tracts, filiform polyps, and fistulas in the esophagus.

OTHER ESOPHAGEAL DISORDERS

DIVERTICULA

Diverticula are outpouchings of the wall of the esophagus. A *Zenker's diverticulum* appears in the natural zone of weakness in the posterior hypopharyngeal wall (Killian's triangle) and causes halitosis and regurgitation of saliva and food particles consumed several days previously. When it becomes large and filled with food, such a diverticulum can compress the esophagus and cause dysphagia or complete obstruction. Nasogastric intubation and endoscopy should be performed with utmost care in these patients, since they may cause perforation of the diverticulum. A *midesophageal diverticulum* may be caused by traction from old adhesions or by propulsion associated with esophageal motor abnormalities. An *epiphrenic diverticulum* may be associated with achalasia. Small or medium-sized diverticula and midesophageal and epiphrenic diverticula are usually asymptomatic. *Diffuse intramural diverticulosis* of the esophagus is due to dilation of the deep esophageal glands and may lead to chronic candidiasis or to the development of a stricture high up in the esophagus. These patients may present with dysphagia. Symptomatic Zenker's diverticula are treated by cricopharyngeal myotomy with or without diverticulectomy. Very large symptomatic esophageal diverticula are removed surgically. When they are associated with motor abnormalities, distal myotomy is performed. Strictures associated with diffuse intramural diverticulosis are treated with rubber dilators.

WEBS AND RINGS

Weblike constrictions of the esophagus are usually congenital or inflammatory in origin. Asymptomatic hypopharyngeal webs are demonstrated in <10% of normal individuals. When concentric, they cause intermittent dysphagia to solids. The combination of symptomatic hypopharyngeal webs and iron-deficiency anemia in middle-aged women constitutes *Plummer-Vinson syndrome*. The clinical importance of this syndrome is uncertain. Midesophageal webs are rare. A *lower esophageal mucosal ring* (Schatzki ring) is a thin, weblike constriction located at the squamocolumnar mucosal junction at or near the border of theLES(Fig. 284-2). It invariably produces dysphagia when the lumen diameter is <1.3 cm. Dysphagia to solids is the only symptom, and it is usually episodic. Asymptomatic rings may be present in ~10% of normal individuals. A lower esophageal ring is one of the common causes of dysphagia. Symptomatic webs and mucosal lower esophageal rings are easily treated by dilatation. A *lower esophageal muscular ring* (contractile ring) is located proximal to the site of mucosal rings and may represent an abnormal uppermost segment of the LES. These rings can be recognized by the fact that they are not constant in size and shape. They also may cause dysphagia and should be differentiated from peptic strictures, achalasia, and lower esophageal mucosal ring. Muscular rings do not respond well to dilatation.

HIATAL HERNIA

A *hiatal hernia* is a herniation of part of the stomach into the thoracic cavity through the esophageal hiatus in the diaphragm. A *sliding hiatal hernia* is one in which the gastroesophageal junction and fundus of the stomach slide upward. A sliding hernia may result from weakening of the anchors of the gastroesophageal junction to the diaphragm, from longitudinal contraction of the esophagus, or from increased intraabdominal pressure. Small sliding hernias can be demonstrated commonly during barium studies if intraabdominal pressure is increased. Incidence increases with age; in individuals in the sixth decade of life, the prevalence of such hernias is ~60%. Small sliding hiatal hernias alone probably produce no symptoms but can contribute to reflux esophagitis. A *paraesophageal hernia* is one in which the esophagogastric junction remains fixed in its normal location and a pouch of stomach is herniated beside the gastroesophageal junction through the esophageal hiatus. A paraesophageal or mixed paraesophageal and sliding hernia may become incarcerated and strangulate, leading to acute chest pain, dysphagia, and a mediastinal mass and requiring surgery. A herniated gastric pouch may cause dysphagia, develop gastritis, or ulcerate, causing chronic blood loss. Large paraesophageal hernias should be surgically repaired.

MECHANICAL TRAUMA

Esophageal rupture may be caused by (1) iatrogenic damage from instrumentation of the esophagus or external trauma, (2) increased intraesophageal pressure associated with forceful vomiting or retching (*spontaneous rupture* or *Boerhaave's syndrome*), or (3) diseases of the esophagus such as corrosive esophagitis, esophageal ulcer, and neoplasm. The site of perforation depends on the cause. Instrumental perforation usually occurs in the pharynx or lower esophagus, just above the diaphragm in the posterolateral wall. Esophageal perforation causes severe retrosternal chest pain, which may be worsened by swallowing and breathing. Free air enters the mediastinum and spreads to neighboring structures, causing palpable subcutaneous emphysema in the neck, mediastinal crackling sounds on auscultation, and pneumothorax. With time, secondary infection supervenes, and mediastinal abscess may develop. Esophageal perforation associated with vomiting usually deposits gastric contents in the mediastinum and causes severe mediastinal complications. By contrast, instrumental perforation may be clinically mild and free of severe complications. Spontaneous rupture of the esophagus may mimic myocardial infarction, pancreatitis, or rupture of an abdominal viscus. Symptoms of chest pain may be mild, particularly in the elderly. Mediastinal emphysema may develop late. An x-ray of the chest shows abnormalities in most patients, but computed tomography of the chest is more sensitive in detecting mediastinal air. Fluid from pleural effusions may have a high content of (salivary) amylase. The diagnosis is confirmed by swallow of radiopaque contrast material. Gastrografin is used initially, and if no leak is found, a small amount of thin barium is used to confirm the diagnosis. Treatment includes esophageal and gastric suction and parenteral broad-spectrum antibiotics. Surgical drainage and repair of the laceration should be performed as soon as possible. In patients with terminal carcinoma, surgical repair may not be feasible, and patients with minor instrumental perforation can be treated conservatively. Extensive corrosive damage may require esophageal diversion and excision of the damaged portion.

Mucosal Tear (Mallory-Weiss Syndrome) This tear is usually caused by vomiting,

retching, or vigorous coughing. The tear usually involves the gastric mucosa near the squamocolumnar mucosal junction. Patients present with upper gastrointestinal bleeding, which may be severe. In most patients bleeding ceases spontaneously; continued bleeding may respond to vasopressin therapy or angiographic embolization. Surgery is rarely needed.

Intramural Hematoma Emetogenic injury, particularly in patients with bleeding abnormalities, can cause bleeding between the mucosal and muscle layers of the esophagus. The patients develop sudden dysphagia. The diagnosis is made by barium swallow and computed tomography. Resolution is usually spontaneous.

FOREIGN BODIES

Foreign bodies may lodge in the cervical esophagus just beyond theUES, near the aortic arch, or above theLES. Impaction of a bolus of food, particularly a piece of meat or bread, may occur when the esophageal lumen is narrowed due to stricture, carcinoma, or a lower esophageal ring. Acute impaction causes a complete inability to swallow and severe chest pain. Both foreign bodies and food boluses may be removed endoscopically. Use of a meat tenderizer to facilitate passage of a meat bolus is discouraged because of potential esophageal perforation and aspiration pneumonia.

(Bibliography omitted in Palm version)

Back to Table of Contents

285. PEPTIC ULCER DISEASE AND RELATED DISORDERS - *John Del Valle*

PEPTIC ULCER DISEASE

Burning epigastric pain exacerbated by fasting and improved with meals is a symptom complex associated with peptic ulcer disease (PUD). An *ulcer* is defined as disruption of the mucosal integrity of the stomach and/or duodenum leading to a local defect or excavation due to active inflammation. Ulcers occur within the stomach and/or duodenum and are often chronic in nature. Acid peptic disorders are very common in the United States, with 4 million individuals (new cases and recurrences) affected per year. Lifetime prevalence of PUD in the United States is approximately 12% in men and 10% in women. Moreover, an estimated 15,000 deaths per year occur as a consequence of complicated PUD. The financial impact of these common disorders has been substantial, with an estimated burden on health care costs of >$15 billion per year in the United States.

GASTRIC PHYSIOLOGY

Despite the constant attack on the gastroduodenal mucosa by a host of noxious agents (acid, pepsin, bile acids, pancreatic enzymes, drugs, and bacteria), integrity is maintained by an intricate system that provides mucosal defense and repair.

Gastric Anatomy The gastric epithelial lining consists of rugae that contain microscopic gastric pits, each branching into four or five gastric glands made up of highly specialized epithelial cells. The makeup of gastric glands varies with their anatomic location. Glands within the gastric cardia comprise<5% of the gastric gland area and contain mucous and endocrine cells. The majority of gastric glands (75%) are found within the oxyntic mucosa and contain mucous neck, parietal, chief, endocrine, and enterochromaffin cells (Fig. 285-1). Pyloric glands contain mucous and endocrine cells (including gastrin cells) and are found in the antrum.

The parietal cell, also known as the oxyntic cell, is usually found in the neck, or isthmus, or the oxyntic gland. The resting, or unstimulated, parietal cell has prominent cytoplasmic tubulovesicles and intracellular canaliculi containing short microvilli along its apical surface (Fig. 285-2). H^+, K^+-ATPase is expressed in the tubulovesicle membrane; upon cell stimulation, this membrane, along with apical membranes, transforms into a dense network of apical intracellular canaliculi containing long microvilli. Acid secretion, a process requiring high energy, occurs at the apical canalicular surface. Numerous mitochondria (30 to 40% of total cell volume) generate the energy required for secretion.

Gastroduodenal Mucosal Defense The gastric epithelium is under a constant assault by a series of endogenous noxious factors including HCl, pepsinogen/pepsin, and bile salts. In addition, a steady flow of exogenous substances such as medications, alcohol, and bacteria encounter the gastric mucosa. A highly intricate biologic system is in place to provide defense from mucosal injury and to repair any injury that may occur.

The mucosal defense system can be envisioned as a three-level barrier, composed of preepithelial, epithelial, and subepithelial elements (Fig. 285-3). The first line of defense is a mucus-bicarbonate layer, which serves as a physicochemical barrier to multiple

molecules including hydrogen ions. Mucus is secreted in a regulated fashion by gastroduodenal surface epithelial cells. It consists primarily of water (95%) and a mixture of lipids and glycoproteins. Mucin is the constituent glycoprotein that, in combination with phospholipids (also secreted by gastric mucous cells), forms a hydrophobic surface with fatty acids that extend into the lumen from the cell membrane. The mucous gel functions as a nonstirred water layer impeding diffusion of ions and molecules such as pepsin. Bicarbonate, secreted by surface epithelial cells of the gastroduodenal mucosa into the mucous gel, forms a pH gradient ranging from 1 to 2 at the gastric luminal surface and reaching 6 to 7 along the epithelial cell surface. Bicarbonate secretion is stimulated by calcium, prostaglandins, cholinergic input, and luminal acidification.

Surface epithelial cells provide the next line of defense through several factors, including mucus production, epithelial cell ionic transporters that maintain intracellular pH and bicarbonate production, and intracellular tight junctions. If the preepithelial barrier were breached, gastric epithelial cells bordering a site of injury can migrate to restore a damaged region (*restitution*). This process occurs independent of cell division and requires uninterrupted blood flow and an alkaline pH in the surrounding environment. Several growth factors including epidermal growth factor (EGF), transforming growth factor (TGF)a, and basic fibroblast growth factor (FGF) modulate the process of restitution. Larger defects that are not effectively repaired by restitution require cell proliferation. Epithelial cell regeneration is regulated by prostaglandins and growth factors such as EGF and TGF-a. In tandem with epithelial cell renewal, formation of new vessels (*angiogenesis*) within the injured microvascular bed occurs. Both FGF and vascular endothelial growth factor (VEGF) are important in regulating angiogenesis in the gastric mucosa.

An elaborate microvascular system within the gastric submucosal layer is the key component of the subepithelial defense/repair system. A rich submucosal circulatory bed provides HCO_3^-, which neutralizes the acid generated by parietal cell secretion of HCl. Moreover, this microcirculatory bed provides an adequate supply of micronutrients and oxygen while removing toxic metabolic by-products.

Prostaglandins play a central role in gastric epithelial defense/repair (Fig. 285-4). The gastric mucosa contains abundant levels of prostaglandins. These metabolites of arachidonic acid regulate the release of mucosal bicarbonate and mucus, inhibit parietal cell secretion, and are important in maintaining mucosal blood flow and epithelial cell restitution. Prostaglandins are derived from esterified arachidonic acid, which is formed from phospholipids (cell membrane) by the action of phospholipase A_2. A key enzyme that controls the rate-limiting step in prostaglandin synthesis is cyclooxygenase (COX), which is present in two isoforms (COX-1, COX-2), each having distinct characteristics regarding structure, tissue distribution, and expression. COX-1 is expressed in a host of tissues including the stomach, platelets, kidneys, and endothelial cells. This isoform is expressed in a constitutive manner and plays an important role in maintaining the integrity of renal function, platelet aggregation, and gastrointestinal mucosal integrity. In contrast, the expression of COX-2 is inducible by inflammatory stimuli, and it is expressed in macrophages, leukocytes, fibroblasts, and synovial cells. The beneficial effects of nonsteroidal anti-inflammatory drugs (NSAIDs) on tissue inflammation are due to inhibition of COX-2; the toxicity of these drugs (e.g., gastrointestinal mucosal

ulceration and renal dysfunction) is related to inhibition of the COX-1 isoform. The highly COX-2-selective NSAIDs have the potential to provide the beneficial effect of decreasing tissue inflammation while minimizing toxicity in the gastrointestinal tract (see below).

Physiology of Gastric Secretion Hydrochloric acid and pepsinogen are the two principal gastric secretory products capable of inducing mucosal injury. Acid secretion should be viewed as occurring under basal and stimulated conditions. Basal acid production occurs in a circadian pattern, with highest levels occurring during the night and lowest levels during the morning hours. Cholinergic input via the vagus nerve and histaminergic input from local gastric sources (see below) are the principal contributors to basal acid secretion. Stimulated gastric acid secretion occurs primarily in three phases based on the site where the signal originates (cephalic, gastric, and intestinal). Sight, smell, and taste of food are the components of the cephalic phase, which stimulates gastric secretion via the vagus nerve. The gastric phase is activated once food enters the stomach. This component of secretion is driven by nutrients (amino acids and amines) that directly stimulate the G cell to release gastrin, which in turn activates the parietal cell via direct and indirect mechanisms (see below). Distention of the stomach wall also leads to gastrin release and acid production. The last phase of gastric acid secretion is initiated as food enters the intestine and is mediated by luminal distention and nutrient assimilation. A series of pathways that inhibit gastric acid production are also set into motion during these phases. The gastrointestinal hormone somatostatin is released from endocrine cells found in the gastric mucosa (D cells) in response to HCl. Somatostatin can inhibit acid production by both direct (parietal cell) and indirect mechanisms [decreased histamine release from enterochromaffin-like (ECL) cells and gastrin release from G cells]. Additional neural (central and peripheral) and hormonal (secretin, cholecystokinin) factors play a role in counterbalancing acid secretion. Under physiologic circumstances, these phases are occurring simultaneously.

The acid-secreting parietal cell is located in the oxyntic gland, adjacent to other cellular elements (ECL cell, D cell) important in the gastric secretory process (Fig. 285-5). This unique cell also secretes intrinsic factor. The parietal cell expresses receptors for several stimulants of acid secretion including histamine (H_2), gastrin (cholecystokinin B/gastrin receptor) and acetylcholine (muscarinic, M_3). Each of these are G protein-linked, seven transmembrane-spanning receptors. Binding of histamine to the H_2receptor leads to activation of adenylate cyclase and an increase in cyclic AMP. Activation of the gastrin and muscarinic receptors results in activation of the protein kinase C/phosphoinositide signaling pathway. Each of these signaling pathways in turn regulates a series of downstream kinase cascades, which control the acid-secreting pump, H^+, K^+-ATPase. The discovery that different ligands and their corresponding receptors lead to activation of different signaling pathways explains the potentiation of acid secretion that occurs when histamine and gastrin or acetylcholine are combined. More importantly, this observation explains why blocking one receptor type (H_2) decreases acid secretion stimulated by agents that activate a different pathway (gastrin, acetylcholine). Parietal cells also express receptors for ligands that inhibit acid production (prostaglandins, somatostatin, andEGF).

The enzyme H^+, K^+-ATPase is responsible for generating the large concentration of H^+. It is a membrane-bound protein that consists of two subunits,a and b. The active

catalytic site is found within the a subunit; the function of the b subunit is unclear. This enzyme uses the chemical energy of ATP to transfer H^+ ions from parietal cell cytoplasm to the secretory canaliculi in exchange for K^+. The H^+,K^+-ATPase is located within the secretory canaliculus and in nonsecretory cytoplasmic tubulovesicles. The tubulovesicles are impermeable to K^+, which leads to an inactive pump in this location. The distribution of pumps between the nonsecretory vesicles and the secretory canaliculus varies according to parietal cell activity (Fig. 285-2). Under resting conditions, only 5% of pumps are within the secretory canaliculus, whereas upon parietal cell stimulation, tubulovesicles are immediately transferred to the secretory canalicular membrane, where 60 to 70% of the pumps are activated. Proton pumps are recycled back to the inactive state in cytoplasmic vesicles once parietal cell activation ceases.

The chief cell, found primarily in the gastric fundus, synthesizes and secretes pepsinogen, the inactive precursor of the proteolytic enzyme pepsin. The acid environment within the stomach leads to cleavage of the inactive precursor to pepsin and provides the low pH (<2.0) required for pepsin activity. Pepsin activity is significantly diminished at a pH of 4 and irreversibly inactivated and denatured at a pH of ³7. Many of the secretagogues that stimulate acid secretion also stimulate pepsinogen release. The precise role of pepsin in the pathogenesis ofPUDremains to be established.

PATHOPHYSIOLOGIC BASIS OF PEPTIC ULCER DISEASE

PUDencompasses both gastric and duodenal ulcers. Ulcers are defined as a break in the mucosal surface >5 mm in size, with depth to the submucosa. Duodenal (DU) and gastric ulcers (GU) share many common features in terms of pathogenesis, diagnosis, and treatment, but several factors distinguish them from one another.

Epidemiology

Duodenal Ulcers DUs are estimated to occur in 6 to 15% of the western population. The incidence of DUs declined steadily from 1960 to 1980 and has remained stable since then. The death rates, need for surgery, and physician visits have decreased by>50% over the past 30 years. The reason for the reduction in the frequency of DUs is likely related to the decreasing frequency of *Helicobacter pylori*. Before the discovery of *H. pylori*, the natural history of DUs was typified by frequent recurrences after initial therapy. Eradication of *H. pylori* has greatly reduced these recurrence rates.

Gastric Ulcers GUs tend to occur later in life than duodenal lesions, with a peak incidence reported in the sixth decade. More than half of GUs occur in males and are less common thanDUs, perhaps due to the higher likelihood of GUs being silent and presenting only after a complication develops. Autopsy studies suggest a similar incidence of DUs and GUs.

Pathology

Duodenal Ulcers DUsoccur most often in the first portion of duodenum (>95%), with ~90% located within 3 cm of the pylorus. They are usually £1 cm in diameter but can occasionally reach 3 to 6 cm (giant ulcer). Ulcers are sharply demarcated, with depth at

times reaching the muscularis propria. The base of the ulcer often consists of a zone of eosinophilic necrosis with surrounding fibrosis. Malignant duodenal ulcers are extremely rare.

Gastric Ulcers In contrast toDUs,GUs can represent a malignancy. Benign GUs are most often found distal to the junction between the antrum and the acid secretory mucosa. This junction is variable, but in general the antral mucosa extends about two thirds of the distance of the lesser curvature and one third the way up the greater curvature. Benign GUs are quite rare in the gastric fundus and are histologically similar to DUs. Benign GUs associated with *H. pylori* are associated with antral gastritis. In contrast,NSAID-related GUs are not accompanied by chronic active gastritis but may instead have evidence of a chemical gastropathy.

Pathophysiology It is now clear that *H. pylori* andNSAID-induced injury account for the majority ofDUs. Gastric acid contributes to mucosal injury but does not play a primary role.

Duodenal Ulcers Many acid secretory abnormalities have been described inDUpatients (Table 285-1). Of these, average basal and nocturnal gastric acid secretion appear to be increased in DU patients as compared to control; however, the level of overlap between DU patients and control subjects is substantial. The reason for this altered secretory process is unclear, but *H. pylori* infection may contribute to this finding. Accelerated gastric emptying of liquids has been noted in some DU patients but is not consistently observed; its role in DU formation, if any, is unclear. Bicarbonate secretion is significantly decreased in the duodenal bulb of patients with an active DU as compared to control subjects. *H. pylori* infection may also play a role in this process.

Gastric Ulcer As inDUs, the majority ofGUs can be attributed to either *H. pylori* orNSAID-induced mucosal damage. GUs that occur in the prepyloric area or those in the body associated with a DU or a duodenal scar are similar in pathogenesis to DUs. Gastric acid output (basal and stimulated) tends to be normal or decreased in GU patients. When GUs develop in the presence of minimal acid levels, impairment of mucosal defense factors may be present.

Abnormalities in resting and stimulated pyloric sphincter pressure with a concomitant increase in duodenal gastric reflux have been implicated in someGUpatients. Although bile acids, lysolecithin, and pancreatic enzymes may injure gastric mucosa, a definite role for these in GU pathogenesis has not been established. Delayed gastric emptying of solids has been described in GU patients but has not been reported consistently. The observation that patients who have undergone disruption of the normal pyloric barrier (pyloroplasty, gastroenterostomy) often have superficial gastritis without frank ulceration decreases enthusiasm for duodenal gastric reflux as an explanation for GU pathogenesis.

H. pylori and acid peptic disorders Gastric infection with the bacterium *H. pylori* accounts for the majority ofPUD. This organism also plays a role in the development of gastric mucosal-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. Although the entire genome of *H. pylori* has been sequenced, it is still not clear how this organism, which is in the stomach, causes ulceration in the

duodenum, or whether its eradication will lead to a decrease in gastric cancer.

THE BACTERIUM The bacterium, initially named *Campylobacter pyloridis*, is a gram-negative microaerophilic rod found most commonly in the deeper portions of the mucous gel coating the gastric mucosa or between the mucous layer and the gastric epithelium. It may attach to gastric epithelium but under normal circumstances does not appear to invade cells. It is strategically designed to live within the aggressive environment of the stomach. It is S-shaped (~0.5´ 3 um in size) and contains multiple sheathed flagella. Initially, *H. pylori* resides in the antrum but, over time, migrates towards the more proximal segments of the stomach. The organism is capable of transforming into a coccoid form, which represents a dormant state that may facilitate survival in adverse conditions. The bacterium expresses a host of factors that contribute to its ability to colonize the gastric mucosa and produce mucosal injury. Several of the key bacterial factors include urease (converting urea to NH_3 and water, thus alkalinizing the surrounding acidic environment), catalase, lipase, adhesins, platelet-activating factor, cytotoxin-associated gene protein (Cag A), pic B (induces cytokines), and vacuolating cytotoxin (Vac A). Multiple strains of *H. pylori* exist and are characterized by their ability to express several of these factors (Cag A, Vac A, etc.). It is possible that the different diseases related to *H. pylori* infection can be attributed to different strains of the organism with distinct pathogenic features.

EPIDEMIOLOGY The prevalence of *H. pylori* varies throughout the world and depends to a great extent on the overall standard of living in the region. In developing parts of the world, 80% of the population may be infected by the age of 20. In contrast, in the United States, this organism is rare in childhood. The overall prevalence of *H. pylori* in the United States is ~30%, with individuals born before 1950 having a higher rate of infection than those born later. About 10% of Americans<30 are colonized with the bacteria. This rate of colonization increases with age, with about 50% of individuals age 50 being infected. Factors that predispose to higher colonization rates include poor socioeconomic status and less education. These factors, not race, are responsible for the rate of *H. pylori* infection in blacks and Hispanic Americans being double the rate seen in whites of comparable age. A summary of risk factors for *H. pylori* infection is shown inTable 285-2.

Transmission of *H. pylori* occurs from person to person, following an oral-oral or fecal-oral route. The risk of *H. pylori* infection is declining in developing countries. The rate of infection in the United States has fallen by>50% when compared to 30 years ago.

PATHOPHYSIOLOGY H. pylori infection is virtually always associated with a chronic active gastritis, but only 10 to 15% of infected individuals develop frank peptic ulceration. The basis for this difference is unknown. Initial studies suggested that>90% of all DUs were associated with *H. pylori*, but *H. pylori* is present in only 30 to 60% of individuals withDU and 70% of patients withGU. The pathophysiology of ulcers not associated with *H. pylori* orNSAIDingestion [or the rare Zollinger-Ellison syndrome (ZES)] is unclear.

The particular end result of *H. pylori* infection (gastritis,PUD, gastricMALTlymphoma, gastric cancer) is determined by a complex interplay between bacterial and host factors

(Fig. 285-6).

1. *Bacterial factors*: *H. pylori* is able to facilitate gastric residence, induce mucosal injury, and avoid host defense. Different strains of *H. pylori* produce different virulence factors. A specific region of the bacterial genome, the pathogenicity island, encodes the virulence factors Cag A and pic B. Vac A also contributes to pathogenicity, though it is not encoded within the pathogenicity island. These virulence factors, in conjunction with additional bacterial constituents, can cause mucosal damage. Urease, which allows the bacteria to reside in the acidic stomach, generates NH_3, which can damage epithelial cells. The bacteria produce surface factors that are chemotactic for neutrophils and monocytes, which in turn contribute to epithelial cell injury (see below). *H. pylori* makes proteases and phospholipases that break down the glycoprotein lipid complex of the mucous gel, thus reducing the efficacy of this first line of mucosal defense. *H. pylori* expresses adhesins, which facilitate attachment of the bacteria to gastric epithelial cells. Although lipopolysaccharide (LPS) of gram-negative bacteria often plays an important role in the infection, *H. pylori* LPS has low immunologic activity compared to that of other organisms. It may promote a smoldering chronic inflammation.

2. *Host factors*: The host responds to *H. pylori* infection by mounting an inflammatory response, which contributes to gastric epithelial cell damage without providing immunity against infection. The neutrophil response is strong both in acute and chronic infection. In addition, T lymphocytes and plasma cells are components of the chronic inflammatory infiltrate, supporting the involvement of antigen-specific cellular and humoral responses. A number of cytokines are released from both epithelial and immune modulatory cells in response to *H. pylori* infection including the proinflammatory cytokines tumor necrosis factor (TNF)a, interleukin (IL)1a/b, IL-6, interferon (IFN)g, and granulocyte-macrophage colony stimulating factor. Several chemokines such as IL-8 and growth-regulated oncogene (GRO)a, involved in neutrophil recruitment/activation, and RANTES, which recruits mononuclear cells, have been observed in *H. pylori*-infected mucosa.

The reason for *H. pylori*-mediated duodenal ulceration remains unclear. One potential explanation is that gastric metaplasia in the duodenum ofDUpatients permits *H. pylori* to bind to it and produce local injury secondary to the host response. Another hypothesis is that *H. pylori* antral infection could lead to increased acid production, increased duodenal acid, and mucosal injury. Basal and stimulated [meal, gastrin-releasing peptide (GRP)] gastrin release are increased in *H. pylori*-infected individuals, and somatostatin-secreting D cells may be decreased. *H. pylori* infection might induce increased acid secretion through both direct and indirect actions of *H. pylori* and proinflammatory cytokines (IL-8,TNF, and IL-1) on G, D, and parietal cells (Fig. 285-7). *H. pylori* infection has also been associated with decreased duodenal mucosal bicarbonate production. Data supporting and contradicting each of these interesting theories have been demonstrated. Thus, the mechanism by which *H pylori* infection of the stomach leads to duodenal ulceration remains to be established.

NSAIDs-induced disease

EPIDEMIOLOGY NSAIDsrepresent one of the most commonly used medications in the United States. More than 30 billion over-the-counter tablets and 70 million prescriptions

are sold yearly in the United States alone. The spectrum of NSAID-induced morbidity ranges from nausea and dyspepsia (prevalence reported as high as 50 to 60%) to a serious gastrointestinal complication such as frank peptic ulceration complicated by bleeding or perforation in as many as 3 to 4% of users per year. About 20,000 patients die each year from serious gastrointestinal complications from NSAIDs. Unfortunately, dyspeptic symptoms do not correlate with NSAID-induced pathology. Over 80% of patients with serious NSAID-related complications did not have preceding dyspepsia. In view of the lack of warning signs, it is important to identify patients who are at increased risk for morbidity and mortality related to NSAID usage. A summary of established and possible risk factors is presented in Table 285-3.

PATHOPHYSIOLOGY Prostaglandins play a critical role in maintaining gastroduodenal mucosal integrity and repair. It therefore follows that interruption of prostaglandin synthesis can impair mucosal defense and repair, thus facilitating mucosal injury via a systemic mechanism. A summary of the pathogenetic pathways by which systemically administered NSAIDs may lead to mucosal injury is shown in Fig. 285-8.

Injury to the mucosa also occurs as a result of the topical encounter with NSAIDs. Aspirin and many NSAIDs are weak acids that remain in a nonionized lipophilic form when found within the acid environment of the stomach. Under these conditions, NSAIDs migrate across lipid membranes of epithelial cells, leading to cell injury once trapped intracellularly in an ionized form. Topical NSAIDs can also alter the surface mucous layer, permitting back diffusion of H^+ and pepsin, leading to further epithelial cell damage.

Miscellaneous Pathogenetic Factors in Acid Peptic Disease Cigarette smoking has been implicated in the pathogenesis of PUD. Not only have smokers been found to have ulcers more frequently than do nonsmokers, but smoking appears to decrease healing rates, impair response to therapy, and increase ulcer-related complications such as perforation. The mechanism responsible for increased ulcer diathesis in smokers is unknown. Theories have included altered gastric emptying, decreased proximal duodenal bicarbonate production, and cigarette-induced generation of noxious mucosal free radicals. Acid secretion is *not* abnormal in smokers. Despite these interesting theories, a unifying mechanism for cigarette-induced peptic ulcer diathesis has not been established.

Genetic predisposition has also been considered to play a role in ulcer development. First-degree relatives of DU patients are three times as likely to develop an ulcer; however, the potential role of *H. pylori* infection in contacts is a major consideration. Increased frequency of blood group O and of the nonsecretor status have also been implicated as genetic risk factors for peptic diathesis. However, *H. pylori* preferentially binds to group O antigens. Therefore, the role of genetic predisposition in common PUD has not been established.

Psychological stress has been thought to contribute to PUD, but studies examining the role of psychological factors in its pathogenesis have generated conflicting results. Although PUD is associated with certain personality traits (neuroticism), these same traits are also present in individuals with nonulcer dyspepsia (NUD) and other functional and organic disorders. Although more work in this area is needed, no typical PUD

personality has been found.

Diet has also been thought to play a role in peptic diseases. Certain foods can cause dyspepsia, but no convincing studies indicate an association between ulcer formation and a specific diet. This is also true for beverages containing alcohol and caffeine. Specific chronic disorders have been associated withPUD(Table 285-4).

Multiple factors play a role in the pathogenesis ofPUD. The two predominant causes are *H. pylori* infection andNSAIDingestion. PUD not related to *H. pylori* or NSAIDs may be increasing. Independent of the inciting or injurious agent, peptic ulcers develop as a result of an imbalance between mucosal protection/repair and aggressive factors. Gastric acid plays an essential role in mucosal injury.

CLINICAL FEATURES

History Abdominal pain is common to many gastrointestinal disorders, includingDU andGU, but has a poor predictive value for the presence of either DU or GU. Up to 10% of patients withNSAID-induced mucosal disease can present with a complication (bleeding, perforation, and obstruction) without antecedent symptoms. Despite this poor correlation, a careful history and physical examination are essential components of the approach to a patient suspected of having peptic ulcers.

Epigastric pain described as a burning or gnawing discomfort can be present in bothDU andGU. The discomfort is also described as an ill-defined, aching sensation or as hunger pain. The typical pain pattern in DU occurs 90 min to 3 h after a meal and is frequently relieved by antacids or food. Pain that awakes the patient from sleep (between midnight and 3 A.M.) is the most discriminating symptom, with two-thirds of DU patients describing this complaint. Unfortunately, this symptom is also present in one-third of patients withNUD. The pain pattern in GU patients may be different from that in DU patients, where discomfort may actually be precipitated by food. Nausea and weight loss occur more commonly in GU patients. In the United States, endoscopy detects ulcers in<30% of patients who have dyspepsia. Despite this, 40% of these individuals with typical ulcer symptoms had an ulcer crater, and 40% had gastroduodenitis on endoscopic examination.

The mechanism for development of abdominal pain in ulcer patients is unknown. Several possible explanations include acid-induced activation of chemical receptors in the duodenum, enhanced duodenal sensitivity to bile acids and pepsin, or altered gastroduodenal motility.

Variation in the intensity or distribution of the abdominal pain, as well as the onset of associated symptoms such as nausea and/or vomiting, may be indicative of an ulcer complication. Dyspepsia that becomes constant, is no longer relieved by food or antacids, or radiates to the back may indicate a penetrating ulcer (pancreas). Sudden onset of severe, generalized abdominal pain may indicate perforation. Pain worsening with meals, nausea, and vomiting of undigested food suggest gastric outlet obstruction. Tarry stools or coffee ground emesis indicate bleeding.

Physical Examination Epigastric tenderness is the most frequent finding in patients

withGU orDU. Pain may be found to the right of the midline in 20% of patients. Unfortunately, the predictive value of this finding is rather low. Physical examination is critically important for discovering evidence of ulcer complication. Tachycardia and orthostasis suggest dehydration secondary to vomiting or active gastrointestinal blood loss. A severely tender, boardlike abdomen suggests a perforation. Presence of a succussion splash indicates retained fluid in the stomach, suggesting gastric outlet obstruction.

PUD-Related Complications

Gastrointestinal Bleeding Gastrointestinal bleeding is the most common complication observed in PUD. It occurs in ~15% of patients and more often in individuals>60 years old. The higher incidence in the elderly is likely due to the increased use ofNSAIDsin this group. As many as 20% of patients with ulcer-related hemorrhage bleed without any preceding warning signs or symptoms.

Perforation The second most common ulcer-related complication is perforation, being reported in as many as 6 to 7% ofPUDpatients. As in the case of bleeding, the incidence of perforation in the elderly appears to be increasing secondary to increased use ofNSAIDs. Penetration is a form of perforation in which the ulcer bed tunnels into an adjacent organ.DUs tend to penetrate posteriorly into the pancreas, leading to pancreatitis, whereasGUs tend to penetrate into the left hepatic lobe. Gastrocolic fistulas associated with GUs have also been described.

Gastric Outlet Obstruction Gastric outlet obstruction is the least common ulcer-related complication, occurring in 1 to 2% of patients. A patient may have relative obstruction secondary to ulcer-related inflammation and edema in the peripyloric region. This process often resolves with ulcer healing. A fixed, mechanical obstruction secondary to scar formation in the peripyloric areas is also possible. The latter requires endoscopic (balloon dilation) or surgical intervention. Signs and symptoms relative to mechanical obstruction may develop insidiously. New onset of early satiety, nausea, vomiting, increase of postprandial abdominal pain, and weight loss should make gastric outlet obstruction a possible diagnosis.

Differential Diagnosis The list of gastrointestinal and nongastrointestinal disorders that can mimic ulceration of the stomach or duodenum is quite extensive. The most commonly encountered diagnosis among patients seen for upper abdominal discomfort isNUD. NUD, also known as *functional dyspepsia* or *essential dyspepsia*, refers to a group of heterogeneous disorders typified by upper abdominal pain without the presence of an ulcer. Dyspepsia has been reported to occur in up to 30% of the U.S. population. Up to 60% of patients seeking medical care for dyspepsia have a negative diagnostic evaluation. The etiology of NUD is not established, and the potential role of *H. pylori* in NUD remains controversial.

Several additional disease processes that may present with "ulcer-like" symptoms include proximal gastrointestinal tumors, gastroesophageal reflux, vascular disease, pancreaticobiliary disease (biliary colic, chronic pancreatitis), and gastroduodenal Crohn's disease.

Diagnostic Evaluation In view of the poor predictive value of abdominal pain for the presence of a gastroduodenal ulcer and the multiple disease processes that can mimic this disease, the clinician is often confronted with having to establish the presence of an ulcer. Documentation of an ulcer requires either a radiographic (barium study) or an endoscopic procedure.

Barium studies of the proximal gastrointestinal tract are still commonly used as a first test for documenting an ulcer. The sensitivity of older single-contrast barium meals for detecting aDU is as high as 80%, with a double-contrast study providing detection rates as high as 90%. Sensitivity for detection is decreased in small ulcers (<0.5 cm), presence of previous scarring, or in postoperative patients. A DU appears as a well-demarcated crater, most often seen in the bulb. AGU may represent benign or malignant disease. Typically, a benign GU also appears as a discrete crater with radiating mucosal folds originating from the ulcer margin. Ulcers >3 cm in size or those associated with a mass are more often malignant. Unfortunately, up to 8% of GUs that appear to be benign by radiographic appearance are malignant by endoscopy or surgery. Radiographic studies that show a GU must be followed by endoscopy and biopsy.

Endoscopy provides the most sensitive and specific approach for examining the upper gastrointestinal tract. In addition to permitting direct visualization of the mucosa, endoscopy facilitates photographic documentation of a mucosal defect and tissue biopsy to rule out malignancy (GU) or *H. pylori*. Endoscopic examination is particularly helpful in identifying lesions too small to detect by radiographic examination, for evaluation of atypical radiographic abnormalities, or to determine if an ulcer is a source of blood loss.

Although the methods for diagnosing *H. pylori* are outlined in Chap. 154, a brief summary will be included here (Table 285-5). PyloriTek, a biopsy urease test, has a sensitivity and specificity of >90 to 95%. In the interest of making a diagnosis of *H. pylori* without the need for performing endoscopy, several noninvasive methods for detecting this organism have been developed. Three types of studies routinely used include serologic testing, the^{13}C- or^{14}C-urea breath test, and the fecal *H. pylori* antigen test.

Occasionally, specialized testing such as serum gastrin and gastric acid analysis or sham feeding may be needed in individuals with complicated or refractoryPUD (see "Zollinger-Ellison Syndrome," below). Screening for aspirin orNSAIDS(blood or urine) may also be necessary in refractory, *H. pylori*-negative PUD patients.

TREATMENT

Before the discovery of *H. pylori*, the therapy ofPUDdisease was centered on the old dictum by Schwartz of "no acid, no ulcer." Although acid secretion is still important in the pathogenesis of PUD, eradication of *H. pylori* and therapy/prevention ofNSAID-induced disease is the mainstay. A summary of commonly used drugs for treatment of acid peptic disorders is shown in Table 285-6.

Acid Neutralizing/Inhibitory Drugs

Antacids Before we understood the important role of histamine in stimulating parietal cell activity, neutralization of secreted acid with antacids constituted the main form of therapy for peptic ulcers. They are now rarely, if ever, used as the primary therapeutic agent but instead are often used by patients for symptomatic relief of dyspepsia. The most commonly used agents are mixtures of aluminum hydroxide and magnesium hydroxide. Aluminum hydroxide can produce constipation and phosphate depletion; magnesium hydroxide may cause loose stools. Many of the commonly used antacids (e.g., Maalox, Mylanta) have a combination of both aluminum and magnesium hydroxide in order to avoid these side effects. The magnesium-containing preparation should not be used in chronic renal failure patients because of possible hypermagnesemia, and aluminum may cause chronic neurotoxicity in these patients.

Calcium carbonate and sodium bicarbonate are potent antacids with varying levels of potential problems. The long-term use of calcium carbonate (converts to calcium chloride in the stomach) can lead to milk-alkali syndrome (hypercalcemia, hyperphosphatemia with possible renal calcinosis and progression to renal insufficiency). Sodium bicarbonate may induce systemic alkalosis.

H_2Receptor antagonists Four of these agents are presently available (cimetidine, ranitidine, famotidine, and nizatidine), and their structures share homology with histamine (Fig. 285-9). Although each has different potency, all will significantly inhibit basal and stimulated acid secretion to comparable levels when used at therapeutic doses. Moreover, similar ulcer-healing rates are achieved with each drug when used at the correct dosage. Presently, this class of drug is often used for treatment of active ulcers (4 to 6 weeks) in combination with antibiotics directed at eradicating *H. pylori* (see below).

Cimetidine was the first H_2receptor antagonist used for the treatment of acid peptic disorders. The initial recommended dosing profile for cimetidine was 300 mg four times per day. Subsequent studies have documented the efficacy of using 800 mg at bedtime for treatment of active ulcer, with healing rates approaching 80% at 4 weeks. Cimetidine may have weak antiandrogenic side effects resulting in reversible gynecomastia and impotence, primarily in patients receiving high doses for prolonged periods of time (months to years, as inZES). In view of cimetidine's ability to inhibit cytochrome P450, careful monitoring of drugs such as warfarin, phenytoin, and theophylline is indicated with long-term usage. Other rare reversible adverse effects reported with cimetidine include confusion and elevated levels of serum aminotransferases, creatinine, and serum prolactin. Ranitidine, famotidine, and nizatidine are more potent H_2receptor antagonists than cimetidine. Each can be used once a day at bedtime. Comparable nighttime dosing regimens are ranitidine, 300 mg, famotidine, 40 mg, and nizatidine, 300 mg.

Additional rare, reversible systemic toxicities reported with H_2receptor antagonists include pancytopenia, neutropenia, anemia, and thrombocytopenia, with a prevalence rate varying from 0.01 to 0.2%. Cimetidine and rantidine (to a lesser extent) can bind to hepatic cytochrome P450, whereas the newer agents, famotidine and nizatidine, do not.

Proton pump (H^+,K^+-ATPase) inhibitors Omeprazole, lansoprazole, and the newest additions, rabeprazole and pantoprazole, are substituted benzimidazole derivatives that

covalently bind and irreversibly inhibit H^+,K^+-ATPase. These are the most potent acid inhibitory agents available. Omeprazole and lansoprazole are the proton pump inhibitors (PPIs) that have been used for the longest time. Both are acid labile and are administered as enteric-coated granules in a sustained-release capsule that dissolves within the small intestine at a pH of 6. These agents are lipophilic compounds; upon entering the parietal cell, they are protonated and trapped within the acid environment of the tubulovesicular and canalicular system. These agents potently inhibit all phases of gastric acid secretion. Onset of action is rapid, with a maximum acid inhibitory effect between 2 and 6 h after administration and duration of inhibition lasting up to 72 to 96 h. With repeated daily dosing, progressive acid inhibitory effects are observed, with basal and secretagogue-stimulated acid production being inhibited by >95% after 1 week of therapy. The half-life of PPIs is approximately 18 h, thus it can take between 2 and 5 days for gastric acid secretion to return to normal levels once these drugs have been discontinued. Because the pumps need to be activated for these agents to be effective, their efficacy is maximized if they are administered before a meal (e.g., in the morning before breakfast). Standard dosing for omeprazole and lansoprazole is 20 mg and 30 mg once per day, respectively. Mild to moderate hypergastrinemia has been observed in patients taking these drugs. Carcinoid tumors developed in some animals given the drugs preclinically; however, extensive experience has failed to demonstrate gastric carcinoid tumor development in humans. Serum gastrin levels return to normal levels within 1 to 2 weeks after drug cessation. As with any agent that leads to significant hypochlorhydria, PPIs may interfere with absorption of drugs such as ketoconazole, ampicillin, iron, and digoxin. Hepatic cytochrome P450 can be inhibited by these agents, but the overall clinical significance of this observation is not definitely established. Caution should be taken when using warfarin, diazepam, and phenytoin concomitantly with PPIs.

Cytoprotective Agents

Sucralfate Sucralfate is a complex sucrose salt in which the hydroxyl groups have been substituted by aluminum hydroxide and sulfate. This compound is insoluble in water and becomes a viscous paste within the stomach and duodenum, binding primarily to sites of active ulceration. Sucralfate may act by several mechanisms. In the gastric environment, aluminum hydroxide dissociates, leaving the polar sulfate anion, which can bind to positively charged tissue proteins found within the ulcer bed, and providing a physicochemical barrier impeding further tissue injury by acid and pepsin. Sucralfate may also induce a trophic effect by binding growth factors such asEGF, enhance prostaglandin synthesis, stimulate mucous and bicarbonate secretion, and enhance mucosal defense and repair. Toxicity from this drug is rare, with constipation being the most common one reported (2 to 3%). It should be avoided in patients with chronic renal insufficiency to prevent aluminum-induced neurotoxicity. Hypophosphatemia and gastric bezoar formation have also been rarely reported. Standard dosing of sucralfate is 1 g four times per day.

Bismuth-containing preparations Sir William Osler considered bismuth-containing compounds the drug of choice for treatingPUD. The resurgence in the use of these agents is due to their effect against *H. pylori*. Colloidal bismuth subcitrate (CBS) and bismuth subsalicylate (BSS, Pepto-Bismol) are the most widely used preparations. The mechanism by which these agents induce ulcer healing is unclear. Potential

mechanisms include ulcer coating; prevention of further pepsin/HCl-induced damage; binding of pepsin; and stimulation of prostaglandins, bicarbonate, and mucous secretion. Adverse effects with short-term usage are rare with bismuth compounds. Long-term usage with high doses, especially with the avidly absorbed CBS, may lead to neurotoxicity. These compounds are commonly used as one of the agents in an anti-*H. pylori* regimen (see below).

Prostaglandin analogues In view of their central role in maintaining mucosal integrity and repair, stable prostaglandin analogues were developed for the treatment ofPUD. The prostaglandin E_1derivative misoprostal is the only agent of this class approved by the U.S. Food and Drug Administration for clinical use in the prevention ofNSAID-induced gastroduodenal mucosal injury (see below). The mechanism by which this rapidly absorbed drug provides its therapeutic effect is through enhancement of mucosal defense and repair. Prostaglandin analogues enhance mucous bicarbonate secretion, stimulate mucosal blood flow, and decrease mucosal cell turnover. The most common toxicity noted with this drug is diarrhea (10 to 30% incidence). Other major toxicities include uterine bleeding and contractions; misoprostal is contraindicated in women who may be pregnant, and women of childbearing age must be made clearly aware of this potential drug toxicity. The standard therapeutic dose is 200 ug four times per day.

Miscellaneous drugs A number of drugs aimed at treating acid peptic disorders have been developed over the years. In view of their limited utilization in the United States, if any, they will only be listed briefly. Anticholinergics, designed to inhibit activation of the muscarinic receptor in parietal cells, met with limited success due to their relatively weak acid-inhibiting effect and significant side effects (dry eyes, dry mouth, urinary retention). Tricyclic antidepressants have been suggested by some, but again the toxicity of these agents in comparison to the safe, effective drugs already described, precludes their utility. Finally, the licorice extract carbenoxolone has aldosterone-like side effects with fluid retention and hypokalemia, making it an undesirable therapeutic option.

Therapy of *H. pylori* Extensive effort has been placed into determining who of the many individuals with *H. pylori* infection should be treated. The common conclusion arrived at by multiple consensus conferences (National Institutes of Health Consensus Development, American Digestive Health Foundation International Update Conference, European Maastricht Consensus, and Asia Pacific Consensus Conference) is that *H. pylori* should be eradicated in patients with documentedPUD. This holds true independent of time of presentation (first episode or not), severity of symptoms, presence of confounding factors such as ingestion ofNSAIDs, or whether the ulcer is in remission. Some have advocated treating patients with a history of documented PUD who are found to be *H. pylori*-positive by serology or breath testing. Over half of patients with gastricMALTlymphoma experience complete remission of the tumor in response to *H. pylori* eradication. Treating patients withNUD or to prevent gastric cancer remains controversial.

Multiple drugs have been evaluated in the therapy of *H. pylori*. No single agent is effective in eradicating the organism. Combination therapy for 14 days provides the greatest efficacy. A short-time course administration (7 to 10 days), although attractive,

has not proven as successful as the 14-day regimens. The agents used with the greatest frequency include amoxicillin, metronidazole, tetracycline, clarithromycin, and bismuth compounds.

The physician's goal in treatingPUD is to provide relief of symptoms (pain or dyspepsia), promote ulcer healing, and ultimately prevent ulcer recurrence and complications. The greatest impact of understanding the role of *H. pylori* in peptic disease has been the ability to prevent recurrence of what was often a recurring disease. Documented eradication of *H. pylori* in patients with PUD is associated with a dramatic decrease in ulcer recurrence to 4% (as compared to 59%) inGUpatients and 6% (compared to 67%) inDUpatients. Eradication of the organism may lead to diminished recurrent ulcer bleeding. The impact of its eradication on ulcer perforation is unclear.

Suggested treatment regimens for *H. pylori* are outlined in Table 285-7. Choice of a particular regimen will be influenced by several factors including efficacy, patient tolerance, existing antibiotic resistance, and cost of the drugs. The aim for initial eradication rates should be 85 to 90%. Dual therapy [PPIplus amoxicillin, PPI plus clarithromycin, ranitidine bismuth citrate (Tritec) plus clarithromycin] are not recommended in view of studies demonstrating eradication rates of <80 to 85%. The combination of bismuth, metronidazole, and tetracycline was the first triple regimen found effective against *H. pylori*. The combination of two antibiotics plus either a PPI, H_2blocker, or bismuth compound has comparable success rates. Addition of acid suppression assists in providing early symptom relief and may enhance bacterial eradication.

Triple therapy, although effective, has several drawbacks, including the potential for poor patient compliance and drug-induced side effects. Compliance is being addressed somewhat by simplifying the regimens so that patients can take the medications twice a day. Simpler (dual therapy) and shorter regimens (7 and 10 days) are not as effective as triple therapy for 14 days. Two anti-*H. pylori* regimens are available in prepackaged formulation: Prevpac (lansoprazole, clarithromycin, and amoxicillin) and Helidac (bismuth subsalicylate, tetracycline, and metronidazole). The contents of the Prevpac are to be taken twice per day for 14 days, whereas Helidac constituents are taken four times per day with an antisecretory agent (PPI or H_2blocker), also taken for at least 14 days.

Side effects have been reported in up to 20 to 30% of patients on triple therapy. Bismuth may cause black stools, constipation, or darkening of the tongue. The most feared complication with amoxicillin is pseudomembranous colitis, but this occurs in <1 to 2% of patients. Amoxicillin can also lead to antibiotic-associated diarrhea, nausea, vomiting, skin rash, and allergic reaction. Tetracycline has been reported to cause rashes and very rarely hepatotoxicity and anaphylaxis.

One important concern with treating patients who may not need treatment is the potential for development of antibiotic-resistant strains. The incidence and type of antibiotic-resistant *H. pylori* strains vary worldwide. Strains resistant to metronidazole, clarithromycin, amoxicillin, and tetracycline have been described, with the latter two being uncommon. Antibiotic-resistant strains are the most common cause for treatment failure in compliant patients. Unfortunately, in vitro resistance does not predict outcome

in patients. Culture and sensitivity testing of *H. pylori* is not performed routinely. Although resistance to metronidazole has been found in as many as 30% and 95% of isolates in North America and Asia, respectively, triple therapy is effective in eradicating the organism in >50% of patients infected with a resistant strain.

Failure of *H. pylori* eradication with triple therapy is usually due to infection with a resistant organism. Quadruple therapy (Table 285-7) where clarithromycin is substituted for metronidazole (or vice versa) should be the next step. If eradication is still not achieved in a compliant patient, then culture and sensitivity of the organism should be considered.

Reinfection after successful eradication of *H. pylori* is rare in the United States (<1%/year). If recurrent infection occurs within the first 6 months after completing therapy, the most likely explanation is recrudescence as opposed to reinfection, which occurs later in time.

Therapy ofNSAID-Related Gastric or Duodenal Injury Medical intervention for NSAID-related mucosal injury includes treatment of an active ulcer and prevention of future injury. Recommendations for the treatment and prevention of NSAID-related mucosal injury are in Table 285-8. Ideally the injurious agent should be stopped as the first step in the therapy of an active NSAID-induced ulcer. If that is possible, then treatment with one of the acid inhibitory agents (H_2blockers,PPIs) is indicated. Cessation of NSAIDs is not always possible because of the patient's severe underlying disease. Only PPIs can healGUs orDUs, independent of whether NSAIDs are discontinued.

Prevention ofNSAID-induced ulceration can be accomplished by misoprostol (200 ug qid) or aPPI. High-dose H_2blockers (famotidine, 40 mg bid) have also shown some promise. The use ofCOX-2-selective NSAIDs may also reduce injury to gastric mucosa. Two highly selective COX-2 inhibitors, celecoxib and rofecoxib, are 100 times more selective inhibitors of COX-2 than standard NSAIDs, leading to gastric or duodenal mucosal injury that is comparable to placebo. However, evaluation of possible drug toxicities, such as altered renal function and induction of thrombosis, requires more data.

Approach and Therapy: Summary Controversy continues regarding the best approach to the patient who presents with dyspepsia (Chap. 41). The discovery of *H. pylori* and its role in pathogenesis of ulcers has added a new variable to the equation. Previously, if a patient <50 presented with dyspepsia and without alarming signs or symptoms suggestive of an ulcer complication or malignancy, an empirical therapeutic trial with acid suppression was commonly recommended. Although this approach is practiced by some today, an approach presently gaining approval for the treatment of patients with dyspepsia is outlined in Fig. 285-10. The referral to a gastroenterologist is for the potential need of endoscopy and subsequent evaluation and treatment if the endoscopy is negative.

Once an ulcer (GU orDU) is documented, then the main issue at stake is whether *H. pylori* or anNSAID is involved. With *H. pylori* present, independent of the NSAID status, triple therapy is recommended for 14 days, followed by continued acid-suppressing drugs (H_2receptor antagonist orPPIs) for a total of 4 to 6 weeks. Selection of patients for

documentation of *H. pylori* eradication is an area of some debate. The test of choice for documenting eradication is the urea breath test (UBT). The stool antigen study may also hold promise for this purpose and should certainly be performed if UBT is not available. Serologic testing is not useful for the purpose of documenting eradication since antibody titers fall slowly and often do not become undetectable. Two approaches toward documentation of eradication exist: (1) test for eradication only in individuals with a complicated course or in individuals who are frail or with multisystem disease who would do poorly with an ulcer recurrence, and (2) test all patients for successful eradication. Some recommend that patients with complicated ulcer disease or who are frail should be treated with long-term acid suppression, thus making documentation of *H. pylori* eradication a moot point. In view of this discrepancy in practice, it would be best to discuss with the patient the different options available.

Several issues differentiate the approach to aGUversus aDU. GUs, especially of the body and fundus, have the potential of being malignant. Multiple biopsies of a GU should be taken initially; even if these are negative for neoplasm, repeat endoscopy to document healing at 8 to 12 weeks should be performed, with biopsy if the ulcer is still present. About 70% of GUs eventually found to be malignant undergo significant (usually incomplete) healing.

The majority (>90%) ofGUs andDUs heal with the conventional therapy outlined above. A GU that fails to heal after 12 weeks and a DU that doesn't heal after 8 weeks of therapy should be considered refractory. Once poor compliance and persistent *H. pylori* infection have been excluded,NSAIDuse, either inadvertent or surreptitious, must be excluded. In addition, cigarette smoking must be eliminated. For a GU, malignancy must be meticulously excluded. Next, consideration should be given to a gastric hypersecretory state, which can be excluded with gastric acid analysis. Although a subset of patients have gastric acid hypersecretion of unclear etiology as a contributing factor to refractory ulcers,ZESshould be excluded with a fasting gastrin or secretin stimulation test (see below). More than 90% of refractory ulcers (either DUs or GUs) heal after 8 weeks of treatment with higher doses ofPPI(omeprazole, 40 mg/d). This higher dose is also effective in maintaining remission. Surgical intervention may be a consideration at this point; however, other rare causes of refractory ulcers must be excluded before recommending surgery. Rare etiologies of refractory ulcers that may be diagnosed by gastric or duodenal biopsies include: ischemia, Crohn's disease, amyloidosis, sarcoidosis, lymphoma, eosinophilic gastroenteritis, or infection [cytomegalovirus (CMV), tuberculosis, or syphilis].

Surgical Therapy Surgical intervention inPUD can be viewed as being either elective, for treatment of medically refractory disease, or as urgent/emergent, for the treatment of an ulcer-related complication. Refractory ulcers are an exceedingly rare occurrence. Surgery is more often required for treatment of an ulcer-related complication. Gastrointestinal bleeding (Chap. 44), perforation, and gastric outlet obstruction are the three complications that may require surgical intervention.

Hemorrhage is the most common ulcer-related complication, occurring in ~15 to 25% of patients. Bleeding may occur in any age group but is most often seen in older patients (sixth decade or beyond). The majority of patients stop bleeding spontaneously, but in some, endoscopic therapy (Chap. 283) is necessary. Patients unresponsive or

refractory to endoscopic intervention will require surgery (~5% of transfusion-requiring patients).

Free peritoneal perforation occurs in ~2 to 3% ofDUpatients. As in the case of bleeding, up to 10% of these patients will not have antecedent ulcer symptoms. Concomitant bleeding may occur in up to 10% of patients with perforation, with mortality being increased substantially. Peptic ulcer can also penetrate into adjacent organs, especially with a posterior DU, which can penetrate into the pancreas, colon, liver, or biliary tree.

Pyloric channel ulcers orDUs can lead to gastric outlet obstruction in ~2 to 3% of patients. This can result from chronic scarring or from impaired motility due to inflammation and/or edema with pylorospasm. Patients may present with early satiety, nausea, vomiting of undigested food, and weight loss. Conservative management with nasogastric suction, intravenous hydration/nutrition, and antisecretory agents is indicated for 7 to 10 days with the hope that a functional obstruction will reverse. If a mechanical obstruction persists, endoscopic intervention with balloon dilation may be effective. Surgery should be considered if all else fails.

Specific Operations for Duodenal Ulcers Surgical treatment is designed to decrease gastric acid secretion. Operations most commonly performed include vagotomy and drainage (by pyloroplasty, gastroduodenostomy, or gastrojejunostomy), highly selective vagotomy (which does not require a drainage procedure), and vagotomy with antrectomy. The specific procedure performed is dictated by the underlying circumstances: elective vs. emergency, the degree and extent of duodenal ulceration, and the expertise of the surgeon.

Vagotomy is a component of each of these procedures and is aimed at decreasing acid secretion through ablating cholinergic input to the stomach. Unfortunately, both truncal and selective vagotomy (preserves the celiac and hepatic branches) result in gastric atony despite successful reduction of both basal acid output (BAO, decreased by 85%) and maximal acid output (MAO, decreased by 50%). Drainage procedure through pyloroplasty or gastroduodenostomy is required in an effort to compensate for the vagotomy-induced gastric motility disorder. To minimize gastric dysmotility, highly selective vagotomy (also known as parietal cell, super selective, and proximal vagotomy) was developed. Only the vagal fibers innervating the portion of the stomach that contains parietal cells is transected, thus leaving fibers important for regulating gastric motility intact. Although this procedure leads to an immediate decrease in both BAO and stimulated acid output, acid secretion recovers over time. By the end of the first postoperative year, basal and stimulated acid output are ~30 and 50%, respectively, of preoperative levels. Ulcer recurrence rates are higher with highly selective vagotomy, although the overall complication rates are lower (Table 285-9).

The procedure that provides the lowest rates of ulcer recurrence but has the highest complication rate is vagotomy (truncal or selective) in combination with antrectomy. Antrectomy is aimed at eliminating an additional stimulant of gastric acid secretion, gastrin. Gastrin originates from G cells found in the antrum. Two principal types of reanastomoses are used after antrectomy, gastroduodenostomy (Billroth I) or gastrojejunostomy (Billroth II) (Fig. 285-11). Although Billroth I is often preferred over II, severe duodenal inflammation or scarring may preclude its performance.

Of these procedures, highly selective vagotomy may be the one of choice in the elective setting, except in situations where ulcer recurrence rates are high (prepyloric ulcers and those refractory to H_2therapy). Selection of vagotomy and antrectomy may be more appropriate in these circumstances.

These procedures have been traditionally performed by standard laparotomy. The advent of laparoscopic surgery has led several surgical teams to successfully perform highly selective vagotomy, truncal vagotomy/pyloroplasty, and truncal vagotomy/antrectomy through this approach. An increase in the number of laparoscopic procedures for treatment ofPUD is expected.

Specific Operations for Gastric Ulcers The location and the presence of a concomitantDUdictate the operative procedure performed for aGU. Antrectomy (including the ulcer) with a Billroth I anastomosis is the treatment of choice for an antral ulcer. Vagotomy is performed only if a DU is present. Although ulcer excision with vagotomy and drainage procedure has been proposed, the higher incidence of ulcer recurrence makes this a less desirable approach. Ulcers located near the esophagogastric junction may require a more radical approach, a subtotal gastrectomy with a Roux-en-Y esophagogastrojejunostomy (Csende's procedure). A less aggressive approach including antrectomy, intraoperative ulcer biopsy, and vagotomy (Kelling-Madlener procedure) may be indicated in fragile patients with a high GU. Ulcer recurrence approaches 30% with this procedure.

Surgery-Related Complications Complications seen after surgery forPUD are related primarily to the extent of the anatomical modification performed. Minimal alteration (highly selective vagotomy) is associated with higher rates of ulcer recurrence and less gastrointestinal disturbance. More aggressive surgical procedures have a lower rate of ulcer recurrence but a greater incidence of gastrointestinal dysfunction. Overall, morbidity and mortality related to these procedures are quite low. Morbidity associated with vagotomy and antrectomy or pyloroplasty is £5%, with mortality ~1%. Highly selective vagotomy has lower morbidity and mortality rates of 1 and 0.3%, respectively.

In addition to the potential early consequences of any intraabdominal procedure (bleeding, infection, thromboembolism), gastroparesis, duodenal stump leak, and efferent loop obstruction can be observed.

Recurrent Ulceration The risk of ulcer recurrence is directly related to the procedure performed (Table 285-9). Ulcers that recur after partial gastric resection tend to develop at the anastomosis (stomal or marginal ulcer). Epigastric abdominal pain is the most frequent presenting complaint. Severity and duration of pain tend to be more progressive than observed withDUsbefore surgery.

Ulcers may recur for several reasons including incomplete vagotomy, retained antrum, and, less likely, persistent or recurrent *H. pylori* infection.ZESshould have been excluded preoperatively. More recently, surreptitious use ofNSAIDshas been found to be a reason for recurrent ulcers after surgery, especially if the initial procedure was done for an NSAID-induced ulcer. Once *H. pylori* and NSAIDs have been excluded as etiologic factors, the question of incomplete vagotomy or retained gastric antrum should

be explored. For the latter, fasting plasma gastrin levels should be determined. If elevated, retained antrum or ZES (see below) should be considered. A combination of acid secretory analysis and secretin stimulation (see below) can assist in this differential diagnosis. Incomplete vagotomy can be ruled out by gastric acid analysis coupled with sham feeding. In this test, gastric acid output is measured while the patient sees, smells, and chews a meal (without swallowing). The cephalic phase of gastric secretion, which is mediated by the vagus, is being assessed with this study. An increase in gastric acid output in response to sham feeding is evidence that the vagus nerve is intact.

Medical therapy with H_2blockers will heal postoperative ulceration in 70 to 90% of patients. The efficacy ofPPIs has not been fully assessed in this group, but one may anticipate greater rates of ulcer healing compared to those obtained with H_2blockers. Repeat operation (complete vagotomy, partial gastrectomy) may be required in a small subgroup of patients who have not responded to aggressive medical management.

Afferent Loop Syndromes Two types of afferent loop syndrome can occur in patients who have undergone partial gastric resection with Billroth II anastomosis. The most common of the two is bacterial overgrowth in the afferent limb secondary to stasis. Patients may experience postprandial abdominal pain, bloating, and diarrhea with concomitant malabsorption of fats and vitamin B_{12}. Cases refractory to antibiotics may require surgical revision of the loop. The less common afferent loop syndrome can present with severe abdominal pain and bloating that occur 20 to 60 min after meals. Pain is often followed by nausea and vomiting of bile-containing material. The pain and bloating may improve after emesis. The cause of this clinical picture is theorized to be incomplete drainage of bile and pancreatic secretions from an afferent loop that is partially obstructed. Cases refractory to dietary measures may need surgical revision.

Dumping Syndrome Dumping syndrome consists of a series of vasomotor and gastrointestinal signs and symptoms and occurs in patients who have undergone vagotomy and drainage (especially Billroth procedures). Two phases of dumping, early and late, can occur. Early dumping takes place 15 to 30 min after meals and consists of crampy abdominal discomfort, nausea, diarrhea, belching, tachycardia, palpitations, diaphoresis, light-headedness, and, rarely, syncope. These signs and symptoms arise from the rapid emptying of hyperosmolar gastric contents into the small intestine, resulting in a fluid shift into the gut lumen with plasma volume contraction and acute intestinal distention. Release of vasoactive gastrointestinal hormones (vasoactive intestinal polypeptide, neurotensin, motilin) is also theorized to play a role in early dumping.

The late phase of dumping typically occurs 90 min to 3 h after meals. Vasomotor symptoms (light-headedness, diaphoresis, palpitations, tachycardia, and syncope) predominate during this phase. This component of dumping is thought to be secondary to hypoglycemia from excessive insulin release.

Dumping syndrome is most noticeable after meals rich in simple carbohydrates (especially sucrose) and high osmolarity. Ingestion of large amounts of fluids may also contribute. Up to 50% of postvagotomy and drainage patients will experience dumping syndrome to some degree. Signs and symptoms often improve with time, but a severe

protracted picture can occur in up to 1% of patients.

Dietary modification is the cornerstone of therapy for patients with dumping syndrome. Small, multiple (six) meals devoid of simple carbohydrates coupled with elimination of liquids during meals is important. Antidiarrheals and anticholinergic agents are complimentary to diet. The somatostatin analogue octreotide has been successful in diet refractory cases. This drug is administered subcutaneously (50 ug tid), titrated according to clinical response. Recently a long-acting formulation has become available, but its use in dumping syndrome has not been examined.

Postvagotomy Diarrhea Up to 10% of patients may seek medical attention for the treatment of postvagotomy diarrhea. This complication is most commonly observed after truncal vagotomy. Patients may complain of intermittent diarrhea that occurs typically 1 to 2 h after meals. Occasionally the symptoms may be severe and relentless. This is due to a motility disorder from interruption of the vagal fibers supplying the luminal gut. Other contributing factors may include decreased absorption of nutrients (see below), increased excretion of bile acids, and release of luminal factors that promote secretion. Diphenoxylate or loperamide is often useful in symptom control. The bile salt-binding agent cholestyramine may be helpful in severe cases. Surgical reversal of a 10-cm segment of jejunum may yield a substantial improvement in bowel frequency in a subset of patients.

Bile Reflux Gastropathy A subset of post-partial gastrectomy patients will present with abdominal pain, early satiety, nausea, and vomiting, who have as the only finding mucosal erythema of the gastric remnant. Histologic examination of the gastric mucosa reveals minimal inflammation but the presence of epithelial cell injury. This clinical picture is categorized as bile or alkaline reflux gastropathy/gastritis. Although reflux of bile is implicated as the reason for this disorder, the mechanism is unknown. Prokinetic agents (cisapride, 10 to 20 mg before meals and at bedtime) and cholestyramine have been effective treatments. Cisapride may cause cardiac arrhythmias. Severe refractory symptoms may require using either nuclear scanning with ^{99m}Tc-HIDA, to document reflux, or an alkaline challenge test, where 0.1 *N* NaOH is infused into the stomach in an effort to reproduce the patient's symptoms. Surgical diversion of pancreaticobiliary secretions away from the gastric remnant with a Roux-en-Y gastrojejunostomy consisting of a long (50 to 60 cm) Roux limb has been used in severe cases. Bilious vomiting improves, but early satiety and bloating may persist in up to 50% of patients.

Maldigestion and Malabsorption Weight loss can be observed in up to 60% of patients after partial gastric resection. A significant component of this weight reduction is due to decreased oral intake. However, mild steatorrhea can also develop. Reasons for maldigestion/malabsorption include decreased gastric acid production, rapid gastric emptying, decreased food dispersion in the stomach, reduced luminal bile concentration, reduced pancreatic secretory response to feeding, and rapid intestinal transit.

Decreased serum vitamin B_{12} levels can be observed after partial gastrectomy. This is usually not due to deficiency of intrinsic factor (IF), since a minimal amount of parietal cells (source of IF) are removed during antrectomy. Reduced vitamin B_{12} may be due to competition for the vitamin by bacterial overgrowth or inability to split the vitamin from its

protein-bound source due to hypochlorhydria.

Iron-deficiency anemia may be a consequence of impaired absorption of dietary iron in patients with a Billroth II gastrojejunotomy. Absorption of iron salts is normal in these individuals; thus a favorable response to oral iron supplementation can be anticipated. Folate deficiency with concomitant anemia can also develop in these patients. This deficiency may be secondary to decreased absorption or diminished oral intake.

Malabsorption of vitamin D and calcium resulting in osteoporosis and osteomalacia is common after partial gastrectomy and gastrojejunostomy (Billroth II). Osteomalacia can occur as a late complication in up to 25% of post-partial gastrectomy patients. Bone fractures occur twice as commonly in men after gastric surgery as in a control population. It may take years before x-ray findings demonstrate diminished bone density. Elevated alkaline phosphatase, reduced serum calcium, bone pain, and pathologic fractures may be seen in patients with osteomalacia. The high incidence of these abnormalities in this subgroup of patients justifies treating them with vitamin D and calcium supplementation indefinitely. Therapy is especially important in females.

Gastric Adenocarcinoma The incidence of adenocarcinoma in the gastric stump is increased 15 years after resection. Some have reported a four- to fivefold increase in gastric cancer 20 to 25 years after resection. The pathogenesis is unclear but may involve alkaline reflux, bacterial proliferation, or hypochlorhydria. Endoscopic screening every other year may detect surgically treatable disease.

RELATED CONDITIONS

ZOLLINGER-ELLISON SYNDROME

Severe peptic ulcer diathesis secondary to gastric acid hypersecretion due to unregulated gastrin release from a non-b cell endocrine tumor (gastrinoma) defines the components of theZES. Initially, ZES was typified by aggressive and refractory ulceration in which total gastrectomy provided the only chance for enhancing survival. Today ZES can be cured by surgical resection in up to 30% of patients.

Epidemiology The incidence ofZESvaries from 0.1 to 1% of individuals presenting withPUD. Males are more commonly affected than females, and the majority of patients are diagnosed between ages 30 and 50. Gastrinomas are classified into sporadic tumors (more common) and those associated with multiple endocrine neoplasia (MEN) type I (see below).

Pathophysiology Hypergastremia originating from an autonomous neoplasm is the driving force responsible for the clinical manifestations inZES. Gastrin stimulates acid secretion through gastrin receptors on parietal cells and by inducing histamine release fromECLcells. Gastrin also has a trophic action on gastric epithelial cells. Longstanding hypergastrinemia leads to markedly increased gastric acid secretion through both parietal cell stimulation and increased parietal cell mass. The increased gastric acid output leads to the peptic ulcer diathesis, erosive esophagitis, and diarrhea.

Tumor Distribution Although early studies suggested that the vast majority of

gastrinomas occurred within the pancreas, a significant number of these lesions are extrapancreatic. Over 80% of these tumors are found within the hypothetical gastrinoma triangle (confluence of the cystic and common bile ducts superiorly, junction of the second and third portions of the duodenum inferiorly, and junction of the neck and body of the pancreas medially). Duodenal tumors constitute the most common nonpancreatic lesion; up to 50% of gastrinomas are found here. Less common extrapancreatic sites include stomach, bones, ovaries, heart, liver, and lymph nodes. More than 60% of tumors are considered malignant, with up to 30 to 50% of patients having multiple lesions or metastatic disease at presentation. Histologically, gastrin-producing cells appear well differentiated, expressing markers typically found in endocrine neoplasms (chromogranin, neuron-specific enolase).

Clinical Manifestations Gastric acid hypersecretion is responsible for the signs and symptoms observed in patients withZES. Peptic ulcer is the most common clinical manifestation, occurring in>90% of gastrinoma patients. Initial presentation and ulcer location (duodenal bulb) may be indistinguishable from commonPUD. Clinical situations that should create suspicion of gastrinoma are ulcers in unusual locations (second part of the duodenum and beyond), ulcers refractory to standard medical therapy, ulcer recurrence after acid-reducing surgery, or ulcers presenting with frank complications (bleeding, obstruction, and perforation). Symptoms of esophageal origin are present in up to two-thirds of patients with ZES, with a spectrum ranging from mild esophagitis to frank ulceration with stricture and Barrett's mucosa.

Diarrhea is the next most common clinical manifestation in up to 50% of patients. Although diarrhea often occurs concomitantly with acid peptic disease, it may also occur independent of an ulcer. Etiology of the diarrhea is multifactorial, resulting from marked volume overload to the small bowel, pancreatic enzyme inactivation by acid, and damage of the intestinal epithelial surface by acid. The epithelial damage can lead to a mild degree of maldigestion and malabsorption of nutrients. The diarrhea may also have a secretory component due to the direct stimulatory effect of gastrin on enterocytes or the cosecretion of additional hormones from the tumor, such as vasoactive intestinal peptide.

Gastrinomas can develop in the presence ofMEN I syndrome (Chap. 93) in approximately 25% of patients. This autosomal dominant disorder involves primarily three organ sites: the parathyroid glands (80 to 90%), pancreas (40 to 80%), and pituitary gland (30 to 60%). The genetic defect in MEN I is in the long arm of chromosome 11 (11q11-q13). In view of the stimulatory effect of calcium on gastric secretion, the hyperparathyroidism and hypercalcemia seen in MEN I patients may have a direct effect on ulcer disease. Resolution of hypercalcemia by parathyroidectomy reduces gastrin and gastric acid output in gastrinoma patients. An additional distinguishing feature inZESpatients with MEN I is the higher incidence of gastric carcinoid tumor development (as compared to patients with sporadic gastrinomas). Gastrinomas tend to be smaller, multiple, and located in the duodenal wall more often than is seen in patients with sporadic ZES. Establishing the diagnosis of MEN I is critical not only from the standpoint of providing genetic counseling to the patient and his or her family but also from the surgical approach recommended.

Diagnosis The first step in the evaluation of a patient suspected of havingZES is to

obtain a fasting gastrin level. A list of clinical scenarios that should arouse suspicion regarding this diagnosis is shown inTable 285-10. Fasting gastrin levels are usually<150 pg/mL. Virtually all gastrinoma patients will have a gastrin level >150 to 200 pg/mL. Measurement of fasting gastrin should be repeated to confirm the clinical suspicion.

Multiple processes can lead to an elevated fasting gastrin level (Table 285-11), with gastric hypochlorhydria or achlorhydria being the most frequent causes. Gastric acid induces feedback inhibition of gastrin release. A decrease in acid production will subsequently lead to failure of the feedback inhibitory pathway, resulting in net hypergastrinemia. Gastrin levels will thus be high in patients using antisecretory agents for the treatment of acid peptic disorders and dyspepsia. *H. pylori* infection can also cause hypergastrinemia.

The next step in establishing a biochemical diagnosis of gastrinoma is to assess acid secretion. Nothing further needs to be done if decreased acid output is observed. In contrast, normal or elevated gastric acid output suggests a need for additional tests. Gastric acid analysis is performed by placing a nasogastric tube in the stomach and drawing samples at 15-min intervals for 1 h during unstimulated or basal state (BAO), followed by continued sampling after administration of intravenous pentagastrin (MAO). Up to 90% of gastrinoma patients may have a BAO of³15 meq/h (normal <4 meq/h). Up to 12% of patients with commonPUD may have comparable levels of acid secretion. A BAO/MAO ratio>0.6 is highly suggestive ofZES, but a ratio <0.6 does not exclude the diagnosis.

Gastrin provocative tests have been developed in an effort to differentiate between the causes of hypergastrinemia and are especially helpful in patients with indeterminant acid secretory studies. The tests are the secretin stimulation test, the calcium infusion study, and a standard meal test. In each of these, a fasted patient has an indwelling intravenous catheter in place for serial blood sampling and an intravenous line in place for secretin or calcium infusion. The patient receives either secretion (intravenous bolus of 2 ug/kg) or calcium (calcium gluconate, 5 mg/kg body weight over 3 h) or is fed a meal. Blood is then drawn at predetermined intervals (10 min and 1 min before and at 2, 5, 10, 15, 20, and 30 min after injection for secretin stimulation and at 30-min intervals during the calcium infusion). The most sensitive and specific gastrin provocative test for the diagnosis of gastrinoma is the secretin study. An increase in gastrin of³200 pg within 15 min of secretin injection has a sensitivity and specificity of >90% forZES. The calcium infusion study is less sensitive and specific than the secretin test, with a rise of >400 pg/mL observed in ~80% of gastrinoma patients. The lower accuracy, coupled with it being a more cumbersome study with greater potential for adverse effects, makes calcium infusion less useful and therefore rarely, if ever, utilized. Rarely, one may observe increasedBAO and hypergastrinemia in a patient who in the past has been categorized as having G cell hyperplasia or hyperfunction. This set of findings may have been due to *H. pylori*. The standard meal test was devised to assist in making the diagnosis of G cell-related hyperactivity, by observing a dramatic increase in gastrin after a meal (>200%). This test is not useful in differentiating between G cell hyperfunction and ZES.

Tumor Localization Once the biochemical diagnosis of gastrinoma has been confirmed, the tumor must be located. Multiple imaging studies have been utilized in an

effort to enhance tumor localization (Table 285-12). The broad range of sensitivity is due to the variable success rates achieved by the different investigative groups. Endoscopic ultrasound (EUS) permits imaging of the pancreas with a high degree of resolution (<5 mm). This modality is particularly helpful in excluding small neoplasms within the pancreas and in assessing the presence of surrounding lymph nodes and vascular involvement. Several types of endocrine tumors express cell-surface receptors for somatostatin. This permits the localization of gastrinomas by measuring the uptake of the stable somatostatin analogue,^{111}In-pentriotide (*octreoscan*) with sensitivity and specificity rates of>75%.

Up to 50% of patients have metastatic disease at diagnosis. Success in controlling gastric acid hypersecretion has shifted the emphasis of therapy towards providing a surgical cure. Detecting the primary tumor and excluding metastatic disease are critical in view of this paradigm shift. Once a biochemical diagnosis has been confirmed, the patient should first undergo an abdominal computed tomographic scan, magnetic resonance imaging, or octreoscan (depending on availability) to exclude metastatic disease. Once metastatic disease has been excluded, an experienced endocrine surgeon may opt for exploratory laparotomy with intraoperative ultrasound or transillumination. In other centers, careful examination of the peripancreatic area withEUS, accompanied by endoscopic exploration of the duodenum for primary tumors, will be performed before surgery. Selective arterial secretin injection (SASI) may be a useful adjuvant for localizing tumors in a subset of patients.

TREATMENT

Treatment of functional endocrine tumors is directed at ameliorating the signs and symptoms related to hormone overproduction, curative resection of the neoplasm, and attempts to control tumor growth in metastatic disease.

PPIsare the treatment of choice and have decreased the need for total gastrectomy. Initial doses of omeprazole or lansoprazole should be in the range of 60 mg/d. Dosing can be adjusted to achieve aBAO<10 meq/h (at the drug trough) in surgery-naive patients and to <5 meq/h in individuals who have previously undergone an acid-reducing operation. Although the somatostatin analogue has inhibitory effects on gastrin release from receptor-bearing tumors and inhibits gastric acid secretion to some extent, PPIs have the advantage of reducing parietal cell activity to a greater degree.

The ultimate goal of surgery would be to provide a definitive cure. Improved understanding of tumor distribution has led to 10-year disease-free intervals as high as 34% in sporadic gastrinoma patients undergoing surgery. A positive outcome is highly dependent on the experience of the surgical team treating these rare tumors. Surgical therapy of gastrinoma patients withMEN I remains controversial because of the difficulty in rendering these patients disease free with surgery. In contrast to the encouraging postoperative results observed in patients with sporadic disease, only 6% of MEN I patients are disease free 5 years after an operation. Some groups suggest surgery only if a clearly identifiable, nonmetastatic lesion is documented by structural studies. Others advocate a more aggressive approach, where all patients free of hepatic metastasis are explored and all detected tumors in the duodenum are resected; this is followed by enucleation of lesions in the pancreatic head, with a distal pancreatectomy to follow.

The outcome of the two approaches has not been clearly defined.

Therapy of metastatic endocrine tumors in general remains suboptimal; gastrinomas are no exception. A host of medical therapeutic approaches including chemotherapy (streptozotocin, 5-fluorouracil, and doxorubicin),IFN-a, and hepatic artery embolization lead to significant toxicity without a substantial improvement in overall survival. Surgical approaches including debulking surgery and liver transplantation for hepatic metastasis have also produced limited benefit. Therefore, early recognition and surgery are the only chances for curing this disease.

The 5- and 10-year survival rates for gastrinoma patients are 62 to 75% and 47 to 53%, respectively. Individuals with the entire tumor resected or those with a negative laparotomy have 5- and 10-year survival rates>90%. Patients with incompletely resected tumors have 5- and 10-year survival of 43% and 25%, respectively. Patients with hepatic metastasis have <20% survival at 5 years. Favorable prognostic indicators include primary duodenal wall tumors, isolated lymph node tumor, and undetectable tumor upon surgical exploration. Poor prognostic indicators include hepatic metastases or the presence of Cushing's syndrome in a sporadic gastrinoma patient.

STRESS-RELATED MUCOSAL INJURY

Patients suffering from shock, sepsis, massive burns, severe trauma, or head injury can develop acute erosive gastric mucosal changes or frank ulceration with bleeding. Classified as stress-induced gastritis or ulcers, injury is most commonly observed in the acid-producing (fundus and body) portions of the stomach. The most common presentation is gastrointestinal bleeding, which is usually minimal but can occasionally be life-threatening. Respiratory failure requiring mechanical ventilation and underlying coagulopathy are risk factors for bleeding, which tends to occur 48 to 72 h after the acute injury or insult.

Histologically, stress injury does not contain inflammation or *H. pylori*; thus "gastritis" is a misnomer. Although elevated gastric acid secretion may be noted in patients with stress ulceration after head trauma (Cushing's ulcer) and severe burns (Curling's ulcer), mucosal ischemia and breakdown of the normal protective barriers of the stomach also play an important role in the pathogenesis. Acid must contribute to injury in view of the significant drop in bleeding noted when acid inhibitors are used as a prophylactic measure for stress gastritis.

Improvement in the general management of intensive care unit patients has led to a significant decrease in the incidence of gastrointestinal bleeding due to stress ulceration. The estimated decrease in bleeding is from 20 to 30% to<15%. This improvement has led to some debate regarding the need for prophylactic therapy. The limited benefit of medical (endoscopic, angiographic) and surgical therapy in a patient with hemodynamically compromising bleeding associated with stress ulcer/gastritis supports the use of preventive measures in high-risk patients (mechanically ventilated, coagulopathy, multiorgan failure, or severe burns). Maintenance of gastric pH >3.5 with continuous infusion of H_2blockers or liquid antacids administered every 2 to 3 h are viable options. Sucralfate slurry (1 g every 4 to 6 h) has also been successful. If bleeding occurs despite these measures, endoscopy, intraarterial vasopressin, or

embolization are options. If all else fails, then surgery should be considered. Although vagotomy and antrectomy may be used, the better approach would be a total gastrectomy, which has an exceedingly high mortality rate in this setting.

GASTRITIS

The term *gastritis* should be reserved for histologically documented inflammation of the gastric mucosa. Gastritis is *not* the mucosal erythema seen during endoscopy and is *not* interchangeable with "dyspepsia." The etiologic factors leading to gastritis are broad and heterogeneous. Gastritis has been classified based on time course (acute vs. chronic), histologic features, and anatomic distribution or proposed pathogenic mechanism (Table 285-13).

The correlation between the histologic findings of gastritis, the clinical picture of abdominal pain or dyspepsia, and endoscopic findings noted on gross inspection of the gastric mucosa is poor. Therefore, there is no typical clinical manifestation of gastritis.

Acute Gastritis The most common causes of acute gastritis are infectious. Acute infection with *H. pylori* induces gastritis. However, *H. pylori* acute gastritis has not been extensively studied. Reported as presenting with sudden onset of epigastric pain, nausea, and vomiting, limited mucosal histologic studies demonstrate a marked infiltrate of neutrophils with edema and hyperemia. If not treated, this picture will evolve into one of chronic gastritis. Hypochlorhydria lasting for up to 1 year may follow acute *H. pylori* infection.

The highly acidic gastric environment may be one reason why infectious processes of the stomach are rare. Bacterial infection of the stomach or phlegmonous gastritis is a rare potentially life-threatening disorder, characterized by marked and diffuse acute inflammatory infiltrates of the entire gastric wall, at times accompanied by necrosis. Elderly individuals, alcoholics, and AIDS patients may be affected. Potential iatrogenic causes include polypectomy and mucosal injection with India ink. Organisms associated with this entity include streptococci, staphylococci, *Escherichia coli*, *Proteus*, and *Haemophilus*. Failure of supportive measures and antibiotics may result in gastrectomy.

Other types of infectious gastritis may occur in immunocompromised individuals such as AIDS patients. Examples include herpetic (herpes simplex) orCMVgastritis. The histologic finding of intranuclear inclusions would be observed in the latter.

Chronic Gastritis Chronic gastritis is identified histologically by an inflammatory cell infiltrate consisting primarily of lymphocytes and plasma cells, with very scant neutrophil involvement. Distribution of the inflammation may be patchy, initially involving superficial and glandular portions of the gastric mucosa. This picture may progress to more severe glandular destruction, with atrophy and metaplasia. Chronic gastritis has been classified according to histologic characteristics. These include superficial atrophic changes and gastric atrophy.

The early phase of chronic gastritis is *superficial gastritis*. The inflammatory changes are limited to the lamina propria of the surface mucosa, with edema and cellular infiltrates separating intact gastric glands. Additional findings may include decreased

mucus in the mucous cells and decreased mitotic figures in the glandular cells. The next stage is *atrophic gastritis*. The inflammatory infiltrate extends deeper into the mucosa, with progressive distortion and destruction of the glands. The final stage of chronic gastritis is *gastric atrophy*. Glandular structures are lost; there is a paucity of inflammatory infiltrates. Endoscopically the mucosa may be substantially thin, permitting clear visualization of the underlying blood vessels.

Gastric glands may undergo morphologic transformation in chronic gastritis. Intestinal metaplasia denotes the conversion of gastric glands to a small intestinal phenotype with small-bowel mucosal glands containing goblet cells. The metaplastic changes may vary in distribution from patchy to fairly extensive gastric involvement. Intestinal metaplasia is an important predisposing factor for gastric cancer (Chap. 90).

Chronic gastritis is also classified according to the predominant site of involvement. Type A refers to the body-predominant form (autoimmune) and type B is the central-predominant form (*H. pylori*-related). This classification is artificial in view of the difficulty in distinguishing these two entities. The term *AB gastritis* has been used to refer to a mixed antral/body picture.

Type A Gastritis The less common of the two forms involves primarily the fundus and body, with antral sparing. Traditionally, this form of gastritis has been associated with pernicious anemia (Chap. 107) in the presence of circulating antibodies against parietal cells and intrinsic factor; thus it is also called *autoimmune gastritis*. *H. pylori* infection can lead to a similar distribution of gastritis. The characteristics of an autoimmune picture are not always present.

Antibodies to parietal cells have been detected in >90% of patients with pernicious anemia and in up to 50% of patients with type A gastritis. Anti-parietal cell antibodies are cytotoxic for gastric mucous cells. The parietal cell antibody is directed against H^+,K^+-ATPase. T cells are also implicated in the injury pattern of this form of gastritis.

Parietal cell antibodies and atrophic gastritis are observed in family members of patients with pernicious anemia. These antibodies are observed in up to 20% of individuals over age 60 and in ~20% of patients with vitiligo and Addison's disease. About half of patients with pernicious anemia have antibodies to thyroid antigens, and about 30% of patients with thyroid disease have circulating anti-parietal cell antibodies. Anti-intrinsic factor antibodies are more specific than parietal cell antibodies for type A gastritis, being present in ~40% of patients with pernicious anemia. Another parameter consistent with this form of gastritis being autoimmune in origin is the higher incidence of specific familial histocompatibility haplotypes such as HLA-B8 and -DR3.

The parietal cell-containing gastric gland is preferentially targeted in this form of gastritis, and achlorhydria results. Parietal cells are the source of intrinsic factor, lack of which will lead to vitamin B_{12}deficiency and its sequelae (megaloblastic anemia, neurologic dysfunction).

Gastric acid plays an important role in feedback inhibition of gastrin release from G cells. Achlorhydria, coupled with relative sparing of the antral mucosa (site of G cells), leads to hypergastrinemia. Gastrin levels can be markedly elevated (>500 pg/mL) in

patients with pernicious anemia.ECL cell hyperplasia with frank development of gastric carcinoid tumors may result from gastrin trophic effects. The role of gastrin in carcinoid development is confirmed by the observation that antrectomy leads to regression of these lesions. Hypergastrinemia and achlorhydria may also be seen in non-pernicious anemia-associated type A gastritis.

Type B gastritis Type B, or antral-predominant, gastritis is the more common form of chronic gastritis. *H. pylori* infection is the cause of this entity. Although described as "antral-predominant," this is likely a misnomer in view of studies documenting the progression of the inflammatory process towards the body and fundus of infected individuals. The conversion to a pan-gastritis is time-dependent -- estimated to require 15 to 20 years. This form of gastritis increases with age, being present in up to 100% of people over age 70. Histology improves after *H. pylori* eradication. The number of *H. pylori* organisms decreases dramatically with progression to gastric atrophy, and the degree of inflammation correlates with the level of these organisms. Early on, with antral-predominant findings, the quantity of *H. pylori* is highest and a dense chronic inflammatory infiltrate of the lamina propria is noted accompanied by epithelial cell infiltration with polymorphonuclear leukocytes (Fig. 285-12).

Multifocal atrophic gastritis, gastric atrophy with subsequent metaplasia, has been observed in chronic *H. pylori*-induced gastritis. This may ultimately lead to development of gastric adenocarcinoma (Fig. 285-13;Chap. 90). *H. pylori* infection is now considered an independent risk factor for gastric cancer. Worldwide epidemiologic studies have documented a higher incidence of *H. pylori* infection in patients with adenocarcinoma of the stomach as compared to control subjects. Seropositivity for *H. pylori* is associated with a three- to sixfold increased risk of gastric cancer. This risk may be as high as ninefold after adjusting for the inaccuracy of serologic testing in the elderly. The mechanism by which *H. pylori* infection leads to cancer is unknown. However, eradication of *H. pylori* as a general preventative measure for gastric cancer is not recommended.

Infection with *H. pylori* is also associated with development of a low grade B cell lymphoma, gastricMALTlymphoma (Chap. 112). The chronic T cell stimulation caused by the infection leads to production of cytokines that promote the B cell tumor. Tumor growth remains dependent upon the presence of *H. pylori* in that its eradication is often associated with complete regression of the tumor. The tumor may take more than a year to regress after treating the infection. Such patients should be followed byEUSevery 2 to 3 months. If the tumor is stable or decreasing in size, no other therapy is necessary. If the tumor grows, it may have become a high-grade B cell lymphoma. When the tumor becomes a high-grade aggressive lymphoma histologically, it loses responsiveness to *H. pylori* eradication.

TREATMENT

Treatment in chronic gastritis is aimed at the sequelae and not the underlying inflammation. Patients with pernicious anemia will require parenteral vitamin B_{12}supplementation on a long-term basis. Eradication of *H. pylori* is not routinely recommended unlessPUD or a low-gradeMALTlymphoma is present.

Miscellaneous Forms of Gastritis *Lymphocytic gastritis* is characterized histologically by intense infiltration of the surface epithelium with lymphocytes. The infiltrative process is primarily in the body of the stomach and consists of mature T cells and plasmacytes. The etiology of this form of chronic gastritis is unknown. It has been described in patients with celiac sprue, but whether there is a common factor associating these two entities is unknown. No specific symptoms suggest lymphocytic gastritis. A subgroup of patients has thickened folds noted on endoscopy. These folds are often capped by small nodules that contain a central depression or erosion; this form of the disease is called *varioliform gastritis*. *H. pylori* probably plays no significant role in lymphocytic gastritis. Therapy with glucocorticoids or sodium cromoglycate has obtained unclear results.

Marked eosinophilic infiltration involving any layer of the stomach (mucosa, muscularis propria, and serosa) is characteristic of *eosinophilic gastritis*. Affected individuals will often have circulating eosinophilia with clinical manifestation of systemic allergy. Involvement may range from isolated gastric disease to diffuse eosinophilic gastroenteritis. Antral involvement predominates, with prominent edematous folds being observed on endoscopy. These prominent antral folds can lead to outlet obstruction. Patients can present with epigastric discomfort, nausea, and vomiting. Treatment with glucocorticoids has been successful.

Several systemic disorders may be associated with *granulomatous gastritis*. Gastric involvement has been observed in Crohn's disease. Involvement may range from granulomatous infiltrates noted only on gastric biopsies to frank ulceration and stricture formation. Gastric Crohn's disease usually occurs in the presence of small-intestinal disease. Several rare infectious processes can lead to granulomatous gastritis, including histoplasmosis, candidiasis, syphilis, and tuberculosis. Other unusual causes of this form of gastritis include sarcoidosis, idiopathic granulomatous gastritis, and eosinophilic granulomas involving the stomach. Establishing the specific etiologic agent in this form of gastritis can be difficult, at times requiring repeat endoscopy with biopsy and cytology. Occasionally, a surgically obtained full-thickness biopsy of the stomach may be required to exclude malignancy.

MENETRIER'S DISEASE

Menetrier's disease is a rare entity characterized by large, tortuous gastric mucosal folds. The differential diagnosis of large gastric folds includesZES, malignancy, infectious etiologies (CMV, histoplasmosis, syphilis), and infiltrative disorders such as sarcoidosis. The mucosal folds in Menetrier's disease are often most prominent in the body and fundus. Histologically, massive foveolar hyperplasia (hyperplasia of surface and glandular mucous cells) is noted, which replaces most of the chief and parietal cells. This hyperplasia produces the prominent folds observed. The pits of the gastric glands elongate and may become extremely tortuous. Although the lamina propria may contain a mild chronic inflammatory infiltrate, Menetrier's disease is *not* considered a form of gastritis. The etiology of this unusual clinical picture is unknown. Overexpression of growth factors such asTGF-amay be involved in the process.

Epigastric pain at times accompanied by nausea, vomiting, anorexia, and weight loss are signs and symptoms in patients with Menetrier's disease. Occult gastrointestinal

bleeding may occur, but overt bleeding is unusual and, when present, is due to superficial mucosal erosions. Between 20 and 100% of patients (depending on time of presentation) develop a protein-losing gastropathy accompanied by hypoalbuminemia and edema. Gastric acid secretion is usually reduced or absent because of the replacement of parietal cells. Large gastric folds are readily detectable by either radiographic (barium meal) or endoscopic methods. Endoscopy with deep mucosal biopsy (and cytology) is required to establish the diagnosis and exclude the other entities that may present in a similar manner. A nondiagnostic biopsy may lead to a surgically obtained full-thickness biopsy to exclude malignancy.

TREATMENT

Medical therapy with anticholinergic agents, prostaglandins,PPIs, prednisone, and H_2receptor antagonists has obtained varying results. Anticholinergics decrease protein loss. A high-protein diet should be recommended to replace protein loss in patients with hypoalbuminemia. Ulcers should be treated with a standard approach. Severe disease with persistent and substantial protein loss may require total gastrectomy. Subtotal gastrectomy is performed by some; it may be associated with higher morbidity and mortality secondary to the difficulty in obtaining a patent and long-lasting anastomosis between normal and hyperplastic tissues.

ACKNOWLEDGEMENT

The author acknowledges the contribution of material to this chapter by Dr. Lawrence Friedman and Dr. Walter Peterson from their chapter on this subject in the 14th edition and is grateful to Pamela Glazer for typing this manuscript.

(Bibliography omitted in Palm version)

Back to Table of Contents

286. DISORDERS OF ABSORPTION - *Henry J. Binder*

Disorders of absorption represent a broad spectrum of conditions with multiple etiologies and varied clinical manifestations. Almost all of these clinical problems are associated with *diminished* intestinal absorption of one or more dietary nutrients and are often referred to as the *malabsorption syndrome*. This latter term is not ideal as it represents a pathophysiologic state, does *not* provide an etiologic explanation for the underlying problem, and should not be considered an adequate final diagnosis. The only clinical situations in which absorption is *increased* are hemochromatosis and Wilson's disease, in which there is increased absorption of iron and copper, respectively.

Most, but not all, of these clinical conditions are associated with *steatorrhea*, an increase in stool fat excretion of >6% of dietary fat intake. Some disorders of absorption are not associated with steatorrhea: Primary lactase deficiency, which represents a congenital absence of the small intestinal brush border disaccharidase enzyme lactase, is only associated with lactose "malabsorption," and pernicious anemia is associated with a marked decrease in intestinal absorption of cobalamin (vitamin B_{12}) due to an absence of gastric parietal cell intrinsic factor required for cobalamin absorption.

Disorders of absorption must be included in the differential diagnosis of diarrhea (Chap. 42) for several reasons. First, diarrhea is frequently associated with and/or is a consequence of the diminished absorption of one or more dietary nutrients. The diarrhea may be secondary either to the intestinal process that is responsible for the steatorrhea or to steatorrhea per se. Thus, celiac sprue (see below) is associated with both extensive morphologic changes in the small intestinal mucosa and reduced absorption of several dietary nutrients; in contrast, the diarrhea of steatorrhea is the result of the effect of nonabsorbed dietary fatty acids on intestinal, usually colonic, ion transport. For example, oleic acid and ricinoleic acid (a bacterially hydroxylated fatty acid that is also the active ingredient in castor oil, a widely used laxative) induce active colonic Cl secretion, most likely secondary to increasing intracellular Ca. In addition, diarrhea per se may result in mild steatorrhea (<11 g fat excretion while on a 100-g fat diet). Second, as diarrhea is both a symptom and a sign, most patients will indicate that they have diarrhea, not that they have fat malabsorption. Third, many intestinal disorders that have diarrhea as a prominent symptom (e.g., ulcerative colitis, traveler's diarrhea secondary to an enterotoxin produced by *Escherichia coli*) do not necessarily have diminished absorption of any dietary nutrient.

Diarrhea as a *symptom* (i.e., when used by patients to describe their bowel movement pattern) may be either a decrease in stool consistency, an increase in stool volume, an increase in number of bowel movements, or any combination of these three changes. In contrast, diarrhea as a *sign* is a quantitative increase in stool water or weight of >200 to 225 mL, or g per 24 h, when a western-type diet is consumed. Individuals consuming a diet with a higher fiber content may normally have a stool weight of up to 400 g/24 h. Thus, it is essential that the clinician clarify what an individual patient means by diarrhea, especially since 10% of patients referred to gastroenterologists for further evaluation of unexplained diarrhea do not have an increase in stool water when it is determined quantitatively. Such patients may have small, frequent, somewhat loose bowel movements with stool urgency that is indicative of proctitis but do not have an increase in stool weight or volume.

It is also critical to establish whether a patient's diarrhea is secondary to diminished absorption of one or more dietary nutrients, in contrast to diarrhea that is due to small- and/or large-intestinal fluid and electrolyte secretion. The former has often been termed *osmotic diarrhea*, while the latter has been referred to as *secretory diarrhea*. Unfortunately, as there can be both secretory and osmotic elements present simultaneously in the same disorder, this separation is not always precise. Nonetheless, two studies, determination of stool electrolytes, and observation of the effect of a fast on stool output can help make this distinction.

The demonstration of the effect of prolonged (>24 h) fasting on stool output can be very effective in suggesting that a *dietary nutrient* is responsible for the individual's diarrhea. A secretory diarrhea associated with enterotoxin-induced traveler's diarrhea would not be affected by prolonged fasting, as enterotoxin-induced stimulation of intestinal fluid and electrolyte secretion is not altered by eating. In contrast, diarrhea secondary to lactose malabsorption in primary lactase deficiency would undoubtedly cease during a prolonged fast. Thus, a substantial decrease in stool output while fasting during a quantitative stool collection of at least 24 h is presumptive evidence that the diarrhea is related to malabsorption of a dietary nutrient. The persistence of stool output while fasting indicates the likelihood that the diarrhea is secretory and that the cause of diarrhea is *not* due to a dietary nutrient. Either a luminal (e.g., *E. coli* enterotoxin) or circulating (e.g., vasoactive intestinal peptide) secretogogue could be responsible for the patient's diarrhea persisting unaltered during a prolonged fast. The observed effects of fasting can be compared and correlated with stool electrolyte and osmolality determinations.

Measurement of stool electrolytes and osmolality requires the comparison of stool Na^+ and K^+concentrations determined in liquid stool to the stool osmolality to determine the presence or absence of a so-called stool osmotic gap. The following formula is used:

The cation concentrations are doubled to estimate stool anion concentrations. The presence of a significant osmotic gap suggests the presence in stool water of a substance(s) other than Na/K/anions that presumably is responsible for the patient's diarrhea. Originally, stool osmolality was measured, but it is almost invariably greater than the required 290 to 300 mosmol/kg H_2O, reflecting bacterial degradation of nonabsorbed carbohydrate either immediately before defecation or in the stool jar while awaiting chemical analysis, even when the stool is refrigerated. As a result, the stool osmolality should be assumed to be 300 mosmol/kg H_2O. When the calculated difference is >50, an osmotic gap is present, suggesting that the diarrhea is due to a nonabsorbed dietary nutrient, e.g., a fatty acid and/or carbohydrate. When this difference is<25 to 50, it is presumed that a dietary nutrient is not responsible for the diarrhea. Since elements of both osmotic (i.e., malabsorption of a dietary nutrient) and secretory diarrhea may be present simultaneously, this separation at times is less clear-cut at the bedside than when used as a teaching example. Ideally, the presence of an osmotic gap will be associated with a marked decrease in stool output during a prolonged fast, while the absence of an osmotic gap will likely be present in an individual whose stool output had not been reduced substantially during a period of

fasting.

NUTRIENT DIGESTION AND ABSORPTION

The lengths of the small intestine and colon are ~300 cm and ~80 cm, respectively. However, the effective functional surface area is approximately 600-fold greater than that of a hollow tube as a result of the presence of folds, villi (in the small intestine), and microvilli. The functional surface area of the small intestine is somewhat greater than that of a doubles tennis court. In addition to nutrient digestion and absorption, the intestinal epithelia have several other functions:

1. *Barrier and immune defense*. The intestine is exposed to a large number of potential antigens, enteric and invasive microorganisms, and is extremely effective preventing the entry of almost all these agents. The intestinal mucosa also synthesizes and secretes secretory IgA globulin.

2. *Fluid and electrolyte absorption and secretion*. The intestine absorbs approximately 7 to 8 L of fluid daily, comprising dietary fluid intake (1 to 2 L/d) and salivary, gastric, pancreatic, biliary, and intestinal fluid (6 to 7 L/d). The intestine also responds to several stimuli, especially bacteria and bacterial enterotoxins, that induce fluid and electrolyte secretion, often leading to diarrhea (Chap. 131).

3. *Synthesis and secretion of several proteins*. The intestinal mucosa is a major site for the production of proteins, including apolipoproteins.

4. *Production of several bioactive amines and peptides*. The intestine presents one of the largest endocrine organs in the body and produces several amines and peptides that serve as paracrine and hormonal mediators of intestinal function.

The small and large intestine are anatomically distinct in that villi are present in the small intestine but are absent in the colon and functionally distinct in that nutrient digestion and absorption take place in the small intestine but not in the colon. No precise anatomic characteristics separate duodenum, jejunum, and ileum, although certain nutrients are absorbed exclusively in specific areas of the small intestine. However, villus cells in the small intestine (and surface epithelial cells in the colon) and crypt cells have distinct anatomic and functional characteristics. Intestinal epithelial cells are continuously renewed, with new proliferating epithelial cells at the base of the crypt migrating over 48 to 72 h to the tip of the villus (or surface of the colon), where they are well-developed epithelial cells with digestive and absorptive function. This high rate of cell turnover explains the relatively rapid resolution of diarrhea and other digestive tract side effects during chemotherapy as new cells not exposed to these toxic agents are produced. Equally important is the paradigm of separation of villus/surface cell and crypt cell function: digestive hydrolytic enzymes are present primarily in the brush border of villus epithelial cells. Absorptive and secretory functions are also separated, with villus/surface cells largely being the site for absorptive function, while secretory function is present in crypts of both the small and large intestine.

Nutrients, minerals, and vitamins are absorbed by one or more active transport mechanisms. (The mechanisms of intestinal fluid and electrolyte absorption and

secretion are discussed inChap. 42.) Active transport mechanisms are energy-dependent and mediated by membrane transport proteins. These transport processes will result in the *net* movement of a substance against or in the absence of an electrochemical concentration gradient. Intestinal absorption of amino acids and monosaccharides, e.g., glucose, is also a specialized form of active transport -- *secondary active transport*. The movement of these actively transported nutrients against a concentration gradient is Na^+-dependent and is due to a Na^+gradient across the apical membrane. The Na^+gradient is maintained by Na^+,K^+-ATPase, the so-called Na^+ pump located on the basolateral membrane, which extrudes Na^+ and maintains a low intracellular [Na] as well as the Na^+gradient across the apical membrane. As a result, active glucose absorption and glucose-stimulated Na^+absorption require both the apical membrane transport protein, SGLT, and the basolateral Na^+,K^+-ATPase. In addition to glucose absorption being Na^+-dependent, glucose also stimulates Na^+ and fluid absorption, which is the physiologic basis of oral rehydration therapy for the treatment of diarrhea (Chap. 42).

Although the intestinal epithelial cells are crucial mediators of absorption and ion and water flow, the several cell types in the lamina propria (e.g., mast cells, macrophages, myofibroblasts) and the enteric nervous system interact with the epithelium to regulate mucosal cell function. The function of the intestine is the result of the integrated responses of and interactions between both intestinal epithelial cells and intestinal muscle.

ENTEROHEPATIC CIRCULATION OF BILE ACIDS

Bile acids are not present in the diet but are synthesized in the liver by a series of enzymatic steps that also represent cholesterol catabolism. Indeed, interruption of the enterohepatic circulation of bile acids can reduce serum cholesterol levels by 10% before a new steady state is established. Bile acids are either primary or secondary: primary bile acids are synthesized in the liver from cholesterol, and secondary bile acids are synthesized from primary bile acids in the intestine by colonic bacterial enzymes. The two primary bile acids are cholic acid and chenodeoxycholic acid; the two most abundant secondary bile acids are deoxycholic acid and lithocholic acid. Approximately 500 mg bile acids are synthesized in the liver daily, conjugated to either taurine or glycine to form tauro-conjugated or glyco-conjugated bile acids, respectively, that are secreted into the duodenum in bile. The primary functions of bile acids are (1) to promote bile flow, (2) to solubilize cholesterol and phospholipid in the gall bladder by mixed micelle formation, and (3) to enhance dietary lipid digestion and absorption by forming mixed micelles in the proximal small intestine.

Bile acids are primarily absorbed by an active, Na^+-dependent process that is located exclusively in the ileum, though bile acids can also be absorbed to a lesser extent by non-carrier-mediated transport processes in the jejunum, ileum, and colon. Conjugated bile acids that enter the colon are deconjugated by colonic bacterial enzymes to unconjugated bile acids and are rapidly absorbed. Colonic bacterial enzymes also dehydroxylate bile acids to secondary bile acids.

Bile acids absorbed from the intestine return to the liver via the portal vein where they are resecreted (Fig. 286-1). Bile acid synthesis is largely autoregulated by

7a-hydroxylase, the initial enzyme in cholesterol degradation. A decrease in the amount of bile acids returning to the liver from the intestine is associated with an increase in bile acid synthesis/cholesterol catabolism, which helps maintain the bile acid pool size relatively constant. However, there is a relatively limited capacity for an increase in bile acid synthesis -- about two to two and one-half times (see below). The bile acid pool size is approximately 4 g and is circulated via the enterohepatic circulation about twice during each meal, or six to eight times during a 24-h period. A relatively small quantity of bile acids is not absorbed and is excreted in stool daily; this fecal loss is matched by hepatic bile acid synthesis.

Defects in any of the steps of the enterohepatic circulation of bile acids can result in a decrease in duodenal concentration of conjugated bile acids and, as a result, steatorrhea. Thus, steatorrhea can be caused by abnormalities in bile acid synthesis and excretion, their physical state in the intestinal lumen, and reabsorption (Table 286-1).

Synthesis Decreased bile acid synthesis and steatorrhea have been demonstrated in chronic liver disease, but steatorrhea is often not a major component of the illness of these patients.

Secretion Although bile acid secretion may be reduced or absent in biliary obstruction, steatorrhea is rarely a significant medical problem in these patients. In contrast, primary biliary cirrhosis represents a defect in canalicular excretion of organic anions, including bile acids, and not infrequently is associated with steatorrhea and its consequences, e.g., chronic bone disease. Thus, the osteomalacia and other chronic bone abnormalities often present in patients with primary biliary cirrhosis and other cholestatic syndromes are secondary to steatorrhea that then leads to calcium and vitamin D malabsorption.

Maintenance of Conjugated Bile Acids In bacterial overgrowth syndromes associated with diarrhea, steatorrhea, and macrocytic anemia, there is an increase in a colonic-type of bacterial flora in the small intestine. The steatorrhea is primarily a result of the decrease in conjugated bile acids secondary to their deconjugation by colonic-type bacteria. Two complementary explanations account for the resulting impairment of micelle formation: (1) unconjugated bile acids are rapidly absorbed in the jejunum by nonionic diffusion resulting in a reduced concentration of duodenal bile acids; and (2) the critical micellar concentration (CMC) of unconjugated bile acids is higher than that of conjugated bile acids, and therefore unconjugated bile acids are less effective than conjugated bile acids in micelle formation.

Reabsorption Ileal dysfunction caused by either Crohn's disease or surgical resection results in a decrease in bile acid reabsorption in the ileum and an *increase* in the delivery of bile acids to the large intestine. The resulting clinical consequences -- diarrhea with or without steatorrhea -- is determined by the *degree* of ileal dysfunction and the *response* of the enterohepatic circulation to bile acid losses (Table 286-2). Patients with limited ileal disease or resection will often have diarrhea, but not steatorrhea. The diarrhea, a result of bile acids in the colon stimulating active Cl secretion, has been called *bile acid diarrhea*, or cholorrheic enteropathy, and responds promptly to cholestyramine, an anion-binding resin. Such patients do not develop

steatorrhea because hepatic synthesis of bile acids increases to compensate for the rate of fecal bile acid losses resulting in maintenance of both the bile acid pool size and the intraduodenal concentrations of bile acids. In contrast, patients with greater degrees of ileal disease and/or resection will often have diarrhea and steatorrhea that does not respond to cholestyramine. In this situation, ileal disease is also associated with increased amounts of bile acids entering the colon; however, hepatic synthesis can no longer increase sufficiently to maintain the bile acid pool size. As a consequence, the intraduodenal concentration of bile acids is also reduced to less than theCMC, resulting in impaired micelle formation and steatorrhea. This second situation is often called *fatty acid diarrhea*. Although cholestyramine may not be effective (and may even increase the diarrhea by further depleting the intraduodenal bile acid concentration), a low-fat diet to reduce fatty acids entering the colon can be effective. Two clinical features, the length of ileum removed and the degree of steatorrhea, can predict whether an individual patient will or will not respond to cholestyramine. Unfortunately, these predictors are imperfect, and a therapeutic trial of cholestyramine is often necessary to establish whether an individual patient will benefit from cholestyramine.Table 286-2 contrasts the characteristics of bile acid diarrhea (small ileal dysfunction) and fatty acid diarrhea (large ileal dysfunction).

LIPIDS

Steatorrhea is caused by one or more defects in the digestion and absorption of dietary fat. Average intake of dietary fat in the United States is approximately 120 to 150 g/d, and fat absorption is linear to dietary fat intake. The total load of fat presented to the small intestine is considerably greater, as substantial amounts of lipid are secreted in bile each day. (See above for discussion of enterohepatic circulation of bile acids.) Three types of fatty acids compose fats: long-chain fatty acids (LCFAs), medium-chain fatty acids (MCFAs), and short-chain fatty acids (SCFAs) (Table 286-3). Dietary fat is exclusively composed of long-chain triglycerides (LCTs), i.e., glycerol that is bound via ester-linkages to three LCFAs. While the majority of dietary LCFAs have carbon chain lengths of 16 or 18, fatty acids of carbon chain length >12 are metabolized in the same manner; saturated and unsaturated fatty acids are handled identically.

Assimilation of dietary lipid requires several integrated processes that can be divided into (1) an intraluminal, or digestive, phase; (2) a mucosal, or absorptive, phase; and (3) a delivery, or postabsorptive, phase. An abnormality at any site of this process can cause steatorrhea (Table 286-4). Therefore, it is essential that any patient with steatorrhea be evaluated to identify the specific physiologic defect in overall lipid digestion-absorption as therapy will be determined by the specific cause responsible for the steatorrhea.

The digestive phase has two components, *lipolysis* and *micellar formation*. Although dietary lipid is in the form ofLCTs, the intestinal mucosa does not absorb triglycerides; they must first be hydrolyzed (Fig. 286-2). The initial step in lipid digestion is the formation of emulsions of finely dispersed lipid, which is accomplished by mastication and gastric contractions. Lipolysis, the hydrolysis of triglycerides to free fatty acids, monoglycerides, and glycerol by lipase, is initiated in the stomach by a gastric lipase that has a pH optimum of 4.5 to 6.0. About 20 to 30% of total lipolysis occurs in the stomach. Lipolysis is completed in the duodenum and proximal jejunal by pancreatic

lipase, which is inactivated by pH<7.0. Pancreatic lipolysis is greatly enhanced by the presence of a second pancreatic enzyme, colipase, which facilitates the movement of lipase to the triglyceride.

Impaired lipolysis can lead to steatorrhea and can occur in the presence of pancreatic insufficiency due to chronic pancreatitis in adults or cystic fibrosis in children and adolescents. Normal lipolysis can be maintained by approximately 5% of maximal pancreatic lipase secretion; thus, steatorrhea is a late manifestation of these disorders. A reduction in intraduodenal pH can also result in altered lipolysis as pancreatic lipase is inactivated at pH <7. Thus, ~15% of patients with gastrinoma (Chap. 285) with substantial increases in gastric acid secretion from ectopic production of gastrin (usually from an islet cell adenoma) have diarrhea, and some will have steatorrhea believed secondary to acid-inactivation of pancreatic lipase. Similarly, patients with chronic pancreatitis (who have reduced lipase secretion) often have a decrease in pancreatic bicarbonate secretion, which will also result in a decrease in intraduodenal pH and inactivation of endogenous pancreatic lipase or of therapeutically administered lipase.

Overlying the microvillus membrane of the small intestine is the so-called unstirred water layer, a relatively stagnant aqueous phase that must be traversed by the products of lipolysis that are primarily water-insoluble. Water-soluble mixed micelles provide a mechanism for the water-insoluble products of lipolysis to reach the luminal plasma membrane of villus epithelial cells, the site for lipid absorption. Mixed micelles are molecular aggregates composed of fatty acids, monoglycerides, phospholipids, cholesterol, and conjugated bile acids. Mixed micelles are formed when the concentration of conjugated bile acids is greater than itsCMC, which differs among the several bile acids present in the small intestinal lumen. Conjugated bile acids, synthesized in the liver and excreted into the duodenum in bile, are regulated by the enterohepatic circulation (see above). Steatorrhea can result from impaired movement of fatty acids across the unstirred aqueous fluid layer in two situations: (1) an increase in the relative thickness of the unstirred water layer that occurs in bacterial overgrowth syndromes (see below) secondary to functional stasis (e.g., scleroderma), and (2) a decrease in the *duodenal* concentration of conjugated bile acids below its CMC, resulting in impaired micelle formation. Thus, steatorrhea can be caused by one or more defects in the enterohepatic circulation of bile acids.

Uptake and reesterification represent the *absorptive phase* of lipid digestion-absorption. Although passive diffusion has been thought responsible, a carrier-mediated process may mediate fatty acid and monoglyceride uptake. Regardless of the uptake process, fatty acids and monoglycerides are reesterified by a series of enzymatic steps in the endoplasmic reticulum and Golgi to form triglycerides, the form in which lipid exits from the intestinal epithelial cell. Impaired lipid absorption as a result of either mucosal inflammation (e.g., celiac sprue) and/or intestinal resection can also lead to steatorrhea.

The reesterified triglycerides require the formation of *chylomicrons* to permit their exit from the small-intestinal epithelial cell and their delivery to the liver via the *lymphatics*. Chylomicrons are composed ofb-lipoprotein and contain triglycerides, cholesterol, cholesterol esters, and phospholipids and enter the lymphatics, not the portal vein. Defects in the *postabsorptive phase* of lipid digestion-absorption can also result in steatorrhea, but these disorders are uncommon. Abetalipoproteinemia, or

acanthocytosis, is a rare disorder of impaired synthesis of b-lipoprotein associated with abnormal erythrocytes (acanthocytes), neurologic problems, and steatorrhea. Lipolysis, micelle formation, and lipid uptake are all normal in patients with abetalipoproteinemia, but the reesterified triglyceride cannot exit from the epithelial cell because of the failure to produce chylomicrons. Small-intestinal biopsies of these rare patients in the postprandial state reveal lipid-laden small-intestinal epithelial cells that become perfectly normal in appearance following a 72- to 96-fast. Similarly, abnormalities of intestinal lymphatics (e.g., intestinal lymphangiectasia) may also be associated with steatorrhea as well as protein loss (see below). Steatorrhea can result from defects at any of the several steps in lipid digestion-absorption. The mechanism of lipid digestion-absorption outlined above is limited to *dietary* lipid that is almost exclusively in the form ofLCTs(Table 286-3). Medium-chain triglycerides (MCTs), composed of fatty acids with carbon chain lengths of 8 to 10, are present in large amounts in coconut oil and are used as a nutritional supplement. MCTs can be digested and absorbed by a different pathway from LCTs and at one time held promise as an important treatment of steatorrhea of almost all etiologies. Unfortunately, their therapeutic effects have been less than expected because their use is often not associated with an increase in body weight for reasons that are not completely understood.

MCTs, in contrast toLCTs, do not require pancreatic lipolysis as the triglyceride can be absorbed intact by the intestinal epithelial cell. Further, micelle formation is not necessary for the absorption of MCTs or medium-chain fatty acids, if hydrolyzed by pancreatic lipase. MCTs are absorbed more efficiently than LCTs for the following reasons: (1) the rate of MCT absorption is greater than that of long-chain fatty acids; (2) medium-chain fatty acids following absorption are not reesterified; (3) following absorption, MCTs are hydrolyzed to medium-chain fatty acids; (4) MCTs do not require chylomicron formation for their exit from the intestinal epithelial cells; and (5) their route of exit is via the portal vein and not via lymphatics. Thus, the absorption of MCTs is greater than that of LCTs in pancreatic insufficiency, conditions with reduced intraduodenal bile acid concentrations, small-intestinal mucosal disease, abetalipoproteinemia, and intestinal lymphangiectasia.

SCFAsare not dietary lipids but are synthesized by colonic bacterial enzymes from nonabsorbed carbohydrate and are the anions in highest concentration in stool (between 80 and 130 m*M*). The SCFAs present in stool are primarily acetate, propionate, and butyrate whose carbon chain lengths are 2, 3, and 4, respectively. Butyrate is the primary nutrient for colonic epithelial cells, and its deficiency may be associated with one or more colitides. SCFAs conserve calories and carbohydrate, because carbohydrates not completely absorbed in the small intestine will not be absorbed in the large intestine due to the absence of both disaccharidases and SGLT, the transport protein that mediates monosaccharide absorption. In contrast, SCFAs are rapidly absorbed and stimulate colonic Na-Cl and fluid absorption. Most non-*Clostridium difficile* antibiotic-associated diarrhea is due to antibiotic suppression of colonic microflora, with a resulting decrease in SCFA production. As *C. difficile* accounts for about 10 to 15% of all antibiotic-associated diarrhea, a relative decrease in colonic production of SCFAs is the cause of most antibiotic-associated diarrhea.

The clinical manifestations of steatorrhea are a consequence of both the underlying disorder responsible for the development of steatorrhea and steatorrhea per se.

Depending on the degree of steatorrhea and the level of dietary intake, significant fat malabsorption may lead to weight loss. Steatorrhea per se can be responsible for diarrhea; if the primary cause of the steatorrhea has not been identified, a low-fat diet can often ameliorate the diarrhea by decreasing fecal fat excretion. Steatorrhea is often associated with fat-soluble vitamin deficiency, which will require replacement with water-soluble preparations of these vitamins.

Disorders of absorption may also be associated with malabsorption of other dietary nutrients, most often carbohydrates, with or without a decrease in dietary lipid digestion and absorption. Therefore, knowledge of the mechanism of the digestion and absorption of carbohydrates, proteins, and other minerals and vitamins is useful in the evaluation of patients with altered intestinal nutrient absorption.

CARBOHYDRATES

Carbohydrates in the diet are present in the form of starch, disaccharides (sucrose and lactose), and glucose. Carbohydrates are absorbed only in the small intestine and only in the form of monosaccharides. Therefore, before their absorption, starch and disaccharides must first be digested by pancreatic amylase and intestinal brush border disaccharidases to monosaccharides. Monosaccharide absorption occurs by a Na-dependent process mediated by the brush border transport protein, SGLT.

Lactose malabsorption is the only clinically important disorder of carbohydrate absorption. Lactose, the disaccharide present in milk, requires digestion by brush border lactase to its two constituent monosaccharides, glucose and galactose. Lactase is present in almost all species in the postnatal period but then disappears throughout the animal kingdom, except in humans. Lactase activity persists in many individuals throughout life. Two different types of lactase deficiency exist -- primary and secondary. In *primary lactase deficiency*, a genetically determined decrease or absence of lactase is noted, while all other aspects of both intestinal absorption and brush border enzymes are normal. In a number of non-Caucasian groups, primary lactase deficiency is common in adulthood.Table 286-5 presents the incidence of primary lactase deficiency in several different ethnic groups. Northern European and North American Caucasians are the only population group to maintain small-intestinal lactase activity throughout adult life. It is lactase persistence that is unusual. In contrast, *secondary lactase deficiency* occurs in association with small-intestinal mucosal disease with abnormalities in both structure and function of other brush border enzymes and transport processes. Secondary lactase deficiency is often seen in celiac sprue.

As lactose digestion is rate-limiting compared to glucose/galactose absorption, lactase deficiency is associated with significant lactose malabsorption. Some individuals with lactose malabsorption develop symptoms such as diarrhea, abdominal pain, cramps, and/or flatus. Most individuals with primary lactase deficiency do not have symptoms. Since lactose intolerance may be associated with symptoms suggestive of an irritable bowel syndrome, persistence of such symptoms in an individual with lactose intolerance while on a strict lactose-free diet would suggest that the individual's symptoms were related to irritable bowel syndrome.

Development of symptoms of lactose intolerance is related to several factors:

1. *Amount of lactose in the diet*.

2. *Rate of gastric emptying*. Symptoms are more likely when gastric emptying is rapid than when gastric emptying is slower. Therefore, it is more likely that skim milk will be associated with symptoms of lactose intolerance than whole milk as the rate of gastric emptying following skim milk intake is more rapid. Similarly, the diarrhea observed following subtotal gastrectomy is often a result of lactose intolerance as gastric emptying is accelerated in patients with a gastrojejunostomy.

3. *Small-intestinal transit time*. Although both the small and large intestine contribute to the development of symptoms, many of the symptoms of lactase deficiency are related to the interaction of colonic bacteria and nonabsorbed lactose. More rapid small-intestinal transit makes symptoms more likely.

4. *Colonic compensation by production of*SCFAsfrom nonabsorbed lactose. Reduced levels of colonic microflora, which can occur following antibiotic use, will also be associated with increased symptoms following lactose ingestion, especially in a lactase-deficient individual.

Glucose-galactose or monosaccharide malabsorption may also be associated with diarrhea and is due to a congenital absence of SGLT. Diarrhea is present when these individuals ingest carbohydrates that contain actively transported monosaccharides (e.g., glucose, galactose) but not monosaccharides that are not actively transported (e.g., fructose). Fructose is absorbed by the brush border transport protein, GLUT 5, a facilitated diffusion process that is not Na-dependent and is distinct from SGLT. In contrast, some individuals develop diarrhea as a result of consuming large quantities of sorbitol, a sugar used in diabetic candy; sorbitol is only minimally absorbed due to the absence of an intestinal absorptive transport mechanism for sorbitol.

PROTEINS

Protein is present in food almost exclusively as polypeptides and requires extensive hydrolysis to di- and tripeptides and amino acids before absorption. Proteolysis occurs in both the stomach and small intestine; it is mediated by pepsin secreted as pepsinogen by gastric chief cells and trypsinogen and other peptidases from pancreatic acinar cells. These proenzymes, pepsinogen and trypsinogen, must be activated to pepsin (by pepsin in the presence of a pH< 5) and trypsin (by the intestinal brush border enzyme, enterokinase, and subsequently by trypsin). Proteins are absorbed by separate transport systems for di- and tripeptides and for different types of amino acids, e.g., neutral, dibasic. Alterations in either protein or amino acid digestion and absorption are rarely observed clinically, even in the presence of extensive small-intestinal mucosal inflammation. However, three rare genetic disorders involve protein digestion-absorption: (1) *enterokinase deficiency* is due to an absence of the brush border enzyme that converts the proenzyme trypsinogen to trypsin and is associated with diarrhea, growth retardation, and hypoproteinemia; (2) *Hartnup syndrome*, a defect in neutral amino acid transport, is characterized by a pellagra-like rash and neuropsychiatric symptoms; and (3) *cystinuria*, a defect in dibasic amino acid transport, is associated with renal calculi and chronic pancreatitis.

EVALUATION OF MALABSORPTION

The clues provided by the history, symptoms, and initial preliminary observations will serve to limit extensive, ill-focused, and expensive laboratory and imaging studies. For example, a clinician evaluating a patient with symptoms suggestive of malabsorption who recently had extensive small-intestinal resection for mesenteric ischemia should direct the initial assessment almost exclusively to define whether a short-bowel syndrome might explain the entire clinical picture. Similarly, the development of a pattern of bowel movements suggestive of steatorrhea in a patient with long-standing alcohol abuse and chronic pancreatitis should lead toward assessing pancreatic exocrine function.

The classic picture of malabsorption described in textbooks [3]30 years ago is rarely seen today in most parts of the United States. As a consequence, diseases with malabsorption must be suspected in individuals with less severe symptoms and signs and with subtle evidence of the altered absorption of only a *single* nutrient rather than obvious evidence of the malabsorption of multiple nutrients.

Although diarrhea can be caused by changes in fluid and electrolyte movement in either the small or the large intestine, dietary nutrients are absorbed almost exclusively in the small intestine. Therefore, the demonstration of diminished absorption of a dietary nutrient provides unequivocal evidence of small-intestinal disease, although colonic dysfunction may also be present (e.g., Crohn's disease may involve both small and large intestine). Dietary nutrient absorption may be segmental or heterogeneous along the small intestine and is site-specific. Thus, for example, calcium, iron, and folic acid are exclusively absorbed by active transport processes in the proximal small intestine, especially the duodenum; in contrast, the active transport mechanisms for both cobalamin and bile acids are present only in the ileum. Therefore, in an individual who years previously had had an intestinal resection, the details of which are not presently available, a presentation with evidence of calcium, folic acid, and iron malabsorption but without cobalamin deficiency would make it likely that the duodenum and jejunum, but not ileum, had been resected.

Some nutrients, e.g., glucose, amino acids, and lipids, are absorbed throughout the small intestine, though there is evidence that their rate of absorption is greater in the proximal than in the distal segments. However, following segmental resection of the small intestine, the remaining segments will undergo both morphologic and functional "adaptation" to enhance absorption. Such adaptation is secondary to both the presence of luminal nutrients and hormonal stimuli and may not be complete in humans for several months following the resection. Adaptation is critical for individuals who have undergone massive resection of the small intestine and/or colon to help ensure survival.

Establishing the presence of steatorrhea and identifying its specific cause are often quite difficult for several reasons. Despite attempts to develop tests that do *not* require the collection of stool to document the presence of steatorrhea, the "gold standard" still remains a timed, quantitative stool fat determination. On a practical basis, stool collections are invariably difficult and often incomplete as nobody wants to handle stool. A qualitative test -- Sudan III stain -- has long been available to establish the presence

of an increase in stool fat. This test is rapid and inexpensive but, as a qualitative test, does not establish the degree of fat malabsorption and is best used as a preliminary screening study. Many of the blood, breath, and isotopic tests that have been developed either: (1) do not directly measure fat absorption, (2) have excellent sensitivity when steatorrhea is obvious and severe but have poor sensitivity when steatorrhea is mild, or (3) have not survived the transition from their development in a laboratory to commercial utilization and dissemination.

Despite this situation, the use of routine laboratory studies (i.e., complete blood count, prothrombin time, serum protein determination, alkaline phosphatase) may suggest the presence of dietary nutrient depletion, especially iron, folate, cobalamin, and vitamins D and K. Additional studies include measurement of serum carotene, cholesterol, albumin, iron, folate, and cobalamin levels. The serum carotene level can also be reduced if the patient has poor dietary intake of leafy vegetables.

If steatorrhea and/or altered absorption of other nutrients are suspected, the history, clinical observations, and laboratory testing can help detect deficiency of a dietary nutrient, especially the fat-soluble vitamins A, D, E, or K. Thus, evidence of metabolic bone disease with elevated alkaline phosphatase and/or reduced serum calcium levels would suggest vitamin D malabsorption. A deficiency of vitamin K would be suggested by an elevated prothrombin time in an individual without liver disease who was not taking anticoagulants. Macrocytic anemia would lead to evaluation of whether cobalamin or folic acid malabsorption was present. The presence of iron-deficiency anemia in the absence of occult bleeding from the gastrointestinal tract in either a male or a nonmenstruating female would require evaluation of iron malabsorption and the exclusion of celiac sprue, as iron is absorbed exclusively in the proximal small intestine.

At times, however, a timed (72-h) quantitative stool collection, preferably on a defined diet, must be obtained to determine stool fat content and establish the presence of steatorrhea. The presence of steatorrhea then requires further assessment to establish the pathophysiologic process(es) responsible for the defect in dietary lipid digestion-absorption (Table 286-4). Some of the other studies include the Schilling test, D-xylose test, duodenal mucosal biopsy, small-intestinal radiologic examination, and tests of pancreatic exocrine function.

THE SCHILLING TEST

This test is performed to determine the cause for cobalamin malabsorption. Since cobalamin absorption requires multiple steps, including gastric, pancreatic, and ileal processes, the Schilling test can also be used to assess the integrity of these other organs (Chap. 107). Cobalamin is present primarily in meat. Except in strict vegans, *dietary* cobalamin deficiency is exceedingly uncommon. Dietary cobalamin is bound in the stomach to a glycoprotein called *R-binder protein*, which is synthesized in both the stomach and salivary glands. This cobalamin-R binder complex is formed in the acid milieu of the stomach. Cobalamin absorption has an absolute requirement for intrinsic factor, another glycoprotein synthesized and released by gastric parietal cells, to promote its uptake by specific cobalamin receptors on the brush border of ileal enterocytes. Pancreatic protease enzymes split the cobalamin-R binder complex to release cobalamin in the proximal small intestine, where cobalamin is then bound by

intrinsic factor.

As a consequence, cobalamin absorption may be abnormal in the following:

1. Pernicious anemia, a disease in which immunologically mediated atrophy of gastric parietal cells leads to an absence of both gastric acid and intrinsic factor secretion.

2. Chronic pancreatitis as a result of deficiency of pancreatic proteases to split the cobalamin-R binder complex. Although 50% of patients with chronic pancreatitis have been reported to have an abnormal Schilling test that was corrected by pancreatic enzyme replacement, the presence of a cobalamin-responsive macrocytic anemia in chronic pancreatitis is extremely rare. Although this probably reflects a difference in the digestion/absorption of cobalamin in food versus that in a crystalline form, the Schilling test can still be used to assess pancreatic exocrine function.

3. Achlorhydria or absence of another factor secreted with acid that is responsible for splitting cobalamin away from the proteins in food to which it is bound can lead to vitamin B_{12}malabsorption. Up to one-third of individuals over>60 years have marginal vitamin B_{12}absorption because of the inability to release cobalamin from food; these people have no defects in absorbing crystalline vitamin B_{12}.

4. Bacterial overgrowth syndromes, which are most often secondary to stasis in the small intestine, produce cobalamin deficiency from bacterial utilization of cobalamin (often referred to as *stagnant bowel syndrome*) (see below).

5. Ileal dysfunction (either as a result of inflammation or prior intestinal resection) due to impaired function of the mechanism of cobalamin-intrinsic factor uptake by ileal intestinal epithelial cells.

The Schilling test is performed by administering^{58}Co-labeled cobalamin and collecting urine for 24 h and is dependent upon normal renal and bladder function. Urinary excretion of cobalamin will reflect cobalamin absorption provided that intrahepatic binding sites for cobalamin are fully occupied. To ensure saturation of hepatic cobalamin binding sites so that all absorbed radiolabeled cobalamin will be excreted in urine, 1 mg cobalamin is administered intramuscularly 1 h following ingestion of the radiolabeled cobalamin. The Schilling test may be abnormal (usually defined as<10% excretion in 24 h) in pernicious anemia, chronic pancreatitis, blind loop syndrome, and ileal disease (Table 286-6). Therefore, whenever an abnormal Schilling test is found,^{58}Co-labeled cobalamin should be administered on another occasion either bound to intrinsic factor, with pancreatic enzymes, or following a 5-day course of antibiotics (often tetracycline). A variation of the Schilling test can detect failure to split cobalamin from food proteins. The labeled cobalamin is cooked together with a scrambled egg and administered orally. People with achlorydria will excrete <10% of the labeled cobalamin in the urine. In addition to establishing the etiology for cobalamin deficiency, the Schilling test can be used to help delineate the pathologic process responsible for steatorrhea by assessing ileal, pancreatic, and small-intestinal luminal function.

URINARY D-XYLOSE TEST

THE URINARY D-xylose test for carbohydrate absorption provides an assessment of proximal small-intestinal mucosal function. D-Xylose, a pentose, is absorbed almost exclusively in the proximal small intestine. The D-xylose test is usually performed by giving 25 g D-xylose and collecting urine for 5 h. An abnormal test (<4.5 g excretion) primarily reflects the presence of duodenal/jejunal mucosal disease. The D-xylose test can also be abnormal in patients with blind loop syndrome (as a consequence primarily of abnormal intestinal mucosa) and, as a false-positive study, in patients with large collections of fluid in a third space (i.e., ascites, pleural fluid). The ease of obtaining a mucosal biopsy of the small intestine by endoscopy and the false-negative rate of the D-xylose test have led to its diminished use. When small-intestinal mucosal disease is suspected, a small-intestinal mucosal biopsy should be performed.

RADIOLOGIC EXAMINATION

Radiologic examination of the small intestine using barium contrast (small-bowel series or study) can provide important information in the evaluation of the patient with presumed or suspected malabsorption. These studies are most often performed in conjunction with the examination of the esophagus, stomach, and duodenal bulb, and insufficient barium is given the patient to permit an adequate examination of the small-intestinal mucosa, especially the ileum. As a result, many gastrointestinal radiologists alter the procedure of a barium contrast examination of the small intestine by performing either a small-bowel series in which a large amount of barium is given by mouth without concurrent examination of the esophagus and stomach or an enteroclysis study in which a large amount of barium is introduced into the duodenum via a fluoroscopically placed tube. In addition, many of the diagnostic features initially described by radiologists to denote the presence of small-intestinal disease (e.g., flocculation, segmentation) are rarely seen with current barium suspensions. Nonetheless, in skilled hands barium contrast examination of the small intestine can yield important information. For example, with extensive mucosal disease, dilatation of intestine can be seen as well as dilution of barium from increased intestinal fluid secretion (Fig. 286-3). A normal barium contrast study does *not* exclude the possibility of small-intestinal disease. However, a small-bowel series remains a very useful examination to assess for the presence of anatomic abnormalities, such as strictures and fistulas (as in Crohn's disease) or blind loop syndrome (e.g., multiple jejunal diverticula), and to define the extent of a previous surgical resection.

BIOPSY OF SMALL-INTESTINAL MUCOSA

A small-intestinal mucosal biopsy is essential in the evaluation of a patient with documented steatorrhea or chronic diarrhea (lasting >3 weeks) (Chap. 42). The ready availability of endoscopic equipment to examine the stomach and duodenum has led to their almost uniform use as the preferred method to obtain histologic material of proximal small-intestinal mucosa. The primary indications for a small-intestinal biopsy are (1) evaluation of a patient either with documented or suspected steatorrhea or with chronic diarrhea, and (2) diffuse or focal abnormalities of the small intestine defined on a small-intestinal series. Lesions seen on small-bowel biopsy can be classified into three different categories (Table 286-7): (1) diffuse, specific; (2) patchy, specific; and (3) diffuse, nonspecific.

1. *Diffuse, specific lesions*. There are relatively few diseases associated with altered nutrient absorption that have specific histopathologic abnormalities on small-intestinal mucosal biopsy, and they are uncommon. *Whipple's disease* is characterized by the presence of periodic acid-Schiff (PAS)-positive macrophages in the lamina propria, while the bacilli that are also present may require electron-microscopic examination for identification (Fig. 286-4). *Abetalipoproteinemia* is characterized by a normal mucosal appearance except for the presence of mucosal absorptive cells that contain lipid postprandially and disappear following a prolonged period of either fat-free intake or fasting. *Immune globulin deficiency* is associated with a variety of histopathologic findings on small-intestinal mucosal biopsy. The characteristic feature is the absence or substantial reduction in the number of plasma cells in the lamina propria; the mucosal architecture may be either perfectly normal or flat, i.e., villus atrophy. As patients with immune globulin deficiency are often infected with *Giardia lamblia*, *Giardia* trophozoites may also be seen in the biopsy.

2. *Patchy, specific lesions*. Several diseases are associated with abnormal small-intestinal mucosal biopsies, but the characteristic features that are present have a patchy distribution. As a result, biopsies obtained randomly or in the absence of abnormalities visualized endoscopically may not reveal these diagnostic features. Intestinal *lymphoma* can at times be diagnosed on mucosal biopsy by the identification of malignant lymphoma cells in the lamina propria and submucosa (Chap. 112). The presence of dilated lymphatics in the submucosa and sometimes in the lamina propria indicates the presence of *lymphangiectasia* associated with hypoproteinemia secondary to protein loss into the intestine. *Eosinophilic gastroenteritis* represents a heterogeneous group of disorders with a spectrum of presentations and symptoms with a eosinophilic infiltrate of the lamina propria, with or without peripheral eosinophilia. The patchy nature of the infiltrate as well as its presence in the submucosa often leads to an absence of histopathologic findings on mucosal biopsy. As the involvement of the duodenum in *Crohn's disease* is also submucosal and not necessarily continuous, mucosal biopsies are not the most direct approach to the diagnosis of duodenal Crohn's disease (Chap. 287). Amyloid deposition can be identified by Congo Red stain in some patients with *amyloidosis* involving the duodenum (Chap. 319).

Several microorganisms can be identified on small-intestinal biopsies, establishing a correct diagnosis. Many of these microorganisms are associated with diarrhea that occurs in immunodeficient individuals, especially those with HIV infection, and include *Cryptosporidium*, *Isospora belli*, cytomegalovirus, *Mycobacterium avium intracellulare*, and *G. lamblia*.

3. *Diffuse, nonspecific lesions*. *Celiac sprue* presents with a characteristic mucosal appearance on duodenal/proximal jejunal mucosal biopsy that is not diagnostic of the disease. The diagnosis of celiac sprue is established by clinical, histologic, and immunologic response to a gluten-free diet. *Tropical sprue* is associated with histopathologic findings similar to those of celiac sprue after a tropical or subtropical exposure but does not respond to gluten restriction; most often symptoms improve with antibiotics and folate administration.

Patients with steatorrhea require assessment of *pancreatic exocrine function*, which is often abnormal in chronic pancreatitis. No test assesses pancreatic exocrine function

well. Endoscopic approaches provide excellent assessment of pancreatic duct anatomy but do not assess exocrine function (Chap. 303). One noninvasive study (bentiromide test) of pancreatic exocrine function is based on the feeding of a tripeptide containing *p*-aminobenzoic acid (PABA). Following splitting of PABA by pancreatic proteases, PABA is liberated, absorbed, and excreted in urine. Reduced proteolysis results in reduced urinary excretion of PABA. This test is neither sensitive nor specific.

Table 286-8 summarizes the results of the D-xylose test, Schilling test, and small-intestinal mucosal biopsy in patients with five different causes of steatorrhea.

SPECIFIC DISEASE ENTITIES

CELIAC SPRUE

Celiac sprue is a not uncommon cause of malabsorption of one or more nutrients in Caucasians, especially those of European descent. Celiac sprue has had several other names including nontropical sprue, celiac disease (in children), adult celiac disease, and gluten-sensitive enteropathy. The etiology of celiac sprue is not known, but environmental, genetic, and immunologic factors are important. Celiac sprue has protean manifestations, almost all of which are secondary to nutrient malabsorption, and a varied natural history, with the onset of symptoms occurring at ages ranging from the first year of life through the eighth decade.

The hallmark of celiac sprue is the presence of an abnormal small-intestinal biopsy (Fig. 286-4) and the response of both symptoms, evidence of malabsorption and the histopathologic changes on the small-intestinal biopsy, to the elimination of gluten from the diet. The histopathologic changes have a proximal to distal intestinal distribution of severity, which probably reflects the exposure of the intestinal mucosa to varied amounts of dietary gluten; the degree of symptoms is related to the extent of these histopathologic changes.

The symptoms of celiac sprue may appear with the introduction of cereals in an infant's diet, although there is frequently a spontaneous remission during the second decade of life that may be either permanent or followed by the reappearance of symptoms over several years. Alternatively, the symptoms of celiac sprue may first become evident at almost any age throughout adulthood. In many patients, frequent spontaneous remissions and exacerbations occur. The symptoms range from significant malabsorption of multiple nutrients with diarrhea, steatorrhea, weight loss, and the consequences of nutrient depletion (i.e., anemia and metabolic bone disease) to the absence of any gastrointestinal symptoms but with evidence of the depletion of a single nutrient (e.g., folate deficiency, osteomalacia, edema from protein loss). Asymptomatic relatives of patients with celiac sprue have been identified as having this disease either by small-intestinal biopsy or by serologic studies (e.g., antiendomysial antibodies).

Etiology The etiology of celiac sprue is not known, but environmental, genetic, and immunologic factors all appear to contribute to the disease.

One *environmental* factor is the clear association of the disease with gliadin, a component of gluten that is present in wheat, barley, rice, and, in smaller amounts, oats.

In addition to the role of gluten restriction in treatment, the instillation of gluten into both normal-appearing rectum and distal ileum of patients with celiac sprue results within hours in morphologic changes.

An *immunologic* component to etiology is suspected for three reasons. First, serum antibodies, IgA antigliadin and antiendomysial antibodies, are present, but it is also not known whether such antibodies are primary or secondary to the tissue damage. The antiendomysial antibody has 90 to 95% sensitivity and 90 to 95% specificity, and the antigen recognized by the antiendomysial antibody test is tissue transglutaminase. The relationship of this autoantibody to pathogenetic mechanism(s) responsible for celiac sprue remains to be established. Nonetheless, this antibody will undoubtedly prove extremely useful in establishing the true prevalence of celiac sprue in the general population and may provide important clues to its etiology. Second, treatment with prednisolone for 4 weeks of a patient with celiac sprue who continues to eat gluten will induce a remission and convert the "flat" abnormal duodenal biopsy to a more normal appearing one. Third, gliadin peptides may interact with gliadin-specific T cells that may either mediate tissue injury or induce the release of one or more cytokines that are responsible for the tissue injury.

Genetic factor(s) also appear to be involved in celiac sprue. The incidence of celiac sprue varies widely in different population groups (high in Caucasians, low in blacks and orientals) and is 10% in first-degree relatives of celiac sprue patients. Furthermore, about 95% of patients with celiac sprue express the HLA-DQ2 allele, though only a minority of all persons expressing DQ2 have celiac sprue.

Diagnosis A small-intestinal biopsy is required to establish a diagnosis of celiac sprue (Fig. 286-4). A biopsy should be performed in patients with symptoms and laboratory findings suggestive of nutrient malabsorption and/or deficiency. Since the presentation of celiac sprue is often subtle, without overt evidence of malabsorption or nutrient deficiency, it is important to have a relatively low threshold to perform a biopsy. It is more prudent to perform a biopsy than to obtain another test of intestinal absorption, which can never completely exclude or establish this diagnosis.

The diagnosis of celiac sprue requires the presence of characteristic histopathologic changes on small-intestinal biopsy together with a prompt clinical and histopathologic response following the institution of a gluten-free diet. If serologic studies have detected the presence of IgA antiendomysial antibodies, they too should disappear after a gluten-free diet is started. The changes seen on duodenal/jejunal biopsy are restricted to the mucosa and include: (1) absence or reduced height of villi, resulting in a "flat" appearance; (2) increased loss of villus cells in association with increased crypt cell proliferation resulting in crypt hyperplasia and loss of villus structure, with consequent villus, but not mucosal, atrophy; (3) cuboidal appearance and nuclei that are no longer oriented basally in surface epithelial cells and increased intraepithelial lymphocytes; and (4) increased lymphocytes and plasma cells in the lamina propria. Although these histopathologic features are characteristic of celiac sprue, they are *not* diagnostic because a similar appearance can be seen in tropical sprue, eosinophilic enteritis, and milk-protein intolerance in children and occasionally in lymphoma, bacterial overgrowth, Crohn's disease, and gastrinoma with acid hypersecretion. However, the presence of a characteristic histopathologic appearance that reverts to normal following the initiation of

a gluten-free diet establishes the diagnosis of celiac sprue. Readministration of gluten with or without an additional small-intestinal biopsy is not necessary.

Failure to Respond to Gluten Restriction The most common cause of persistent symptoms in a patient who fulfills all the criteria of the diagnosis of celiac sprue is continued intake of gluten. Gluten is ubiquitous, and significant effort must be made to exclude all gluten from the diet. Use of rice in place of wheat flour is very helpful, and several support groups provide important aid to patients with celiac sprue and to their families. About 90% of patients who have the characteristic findings of celiac sprue will respond to complete dietary gluten restriction. The remainder represent a heterogeneous group (whose condition is often called *refractory sprue*) that includes some patients who (1) respond to restriction of other dietary protein, e.g., soy; (2) respond to glucocorticoids; (3) are "temporary," i.e., the clinical and morphologic findings disappear after several months or years; or (4) fail to respond to all measures and have a fatal outcome, with or without documented complications of celiac sprue.

Mechanism of Diarrhea The diarrhea in celiac sprue has several pathogenetic mechanisms. Diarrhea may be secondary to (1) steatorrhea, which is primarily a result of the changes in jejunal mucosal function; (2) secondary lactase deficiency, a consequence of changes in jejunal brush border enzymatic function; (3) bile acid malabsorption resulting in bile acid-induced fluid secretion in the colon, in cases with more extensive disease involving the ileum; and (4) endogenous fluid secretion resulting from the crypt hyperplasia. Patients with more severe involvement with celiac sprue may obtain temporary improvement with *dietary lactose and fat restriction* while awaiting the full effects of total gluten restriction, which represents primary therapy.

Associated Diseases Celiac sprue is associated with dermatitis herpetiformis (DH), though the association has not been explained. Patients with DH have characteristic papulovesicular lesions that respond to dapsone. Almost all patients with DH have histopathologic changes in the small intestine consistent with celiac sprue, although usually much milder and less diffuse in distribution. Most patients with DH have mild, or no, gastrointestinal symptoms. In contrast, relatively few patients with celiac sprue have DH.

Celiac sprue is also associated with insulin-dependent diabetes mellitus and IgA globulin deficiency. The clinical importance of the former association is that although severe watery diarrhea without evidence of malabsorption is most often seen in patients with "diabetic diarrhea" (Chap. 333), a small-intestinal biopsy must at times be considered to exclude this association.

Complications The most important complication of celiac sprue is the development of a malignancy. An increased incidence of both gastrointestinal and nongastrointestinal neoplasms as well as intestinal lymphoma exists in patients with celiac sprue. For unexplained reasons the occurrence of lymphoma in patients with celiac sprue is higher in Ireland and the United Kingdom than in the United States. The possibility of lymphoma must be considered whenever a patient with celiac sprue previously doing well on a gluten-free diet is no longer responsive to gluten restriction or a patient who presents with clinical and histopathologic features consistent with celiac sprue does not respond to a gluten-free diet. Other complications of celiac sprue include the

development of intestinal ulceration independent of lymphoma and so-called refractory sprue (see above) and collagenous sprue. In *collagenous sprue*, a layer of collagen-like material is present beneath the basement membrane; these patients generally do not respond to a gluten-free diet and often have a poor prognosis.

TROPICAL SPRUE

Tropical sprue is a poorly understood syndrome that affects both expatriates and natives in certain but not all tropical areas and is manifested by chronic diarrhea, steatorrhea, weight loss, and nutritional deficiencies, including those of both folate and cobalamin. This disease affects 5 to 10% of the population in some tropical areas.

Chronic diarrhea in a tropical environment is most often caused by infectious agents including *G. lamblia*, *Yersinia enterocolitica*, *C. difficile*, *Cryptosporidium parvum*, and *Cyclospora cayetanensis*, among other organisms. Tropical sprue should not be entertained as a possible diagnosis until the presence of cysts and trophozoites has been excluded in three stool samples.**Chronic infections of the gastrointestinal tract and diarrhea in patients with or without AIDS are discussed in Chaps. 309 and 131.*

The small-intestinal mucosa in individuals living in tropical areas is not identical to that of individuals who reside in temperate climates. Biopsies reveal a mild alteration of villus architecture with a modest increase in mononuclear cells in the lamina propria, which on occasion can be as severe as that seen in celiac sprue. These changes are observed both in native residents and in expatriates living in tropical regions, are usually associated with mild decreases in absorptive function, but revert to "normal" when an individual moves or returns to a temperate area. Some have suggested that the changes seen in tropical enteropathy and in tropical sprue represent different ends of the spectrum of a single entity, but convincing evidence to support this concept is lacking.

Etiology The etiology of tropical sprue is not known, though because tropical sprue responds to antibiotics, the consensus is that tropical sprue may be caused by one or more infectious agents. Nonetheless, multiple uncertainties regarding the etiology and pathogenesis of tropical sprue exist. First, its occurrence is not evenly distributed in all tropical areas; rather, it is found in specific locations including South India, the Philippines, and several Caribbean islands (e.g., Puerto Rico, Haiti) but is rarely observed in Africa, Jamaica, or Southeast Asia. Second, an occasional individual will not develop symptoms of tropical sprue until long after having left an endemic area. This is the reason why the original term for celiac sprue was *nontropical sprue* to distinguish it from tropical sprue. Third, multiple microorganisms have been identified on jejunal aspirate with relatively little consistency among studies. *Klebsiella pneumoniae*, *Enterobacter cloacae*, or *E. coli* have been implicated in some studies of tropical sprue, while other investigations have favored a role for a toxin produced by one or more of these bacteria. Fourth, the incidence of tropical sprue appears to have decreased substantially during the past decade. One speculation for this reduced occurrence of tropical sprue is the wider use of antibiotics in acute diarrhea especially in travelers to tropical areas from temperate countries. Fifth, the role of folic acid deficiency in the pathogenesis of tropical sprue requires clarification. Folic acid is absorbed exclusively in the duodenum and proximal jejunum, and most patients with tropical sprue have

evidence of folate malabsorption and depletion. Although folate deficiency can cause changes in small-intestinal mucosa that are corrected by folate replacement, the several earlier studies reporting that tropical sprue could be cured by folic acid did not provide an explanation for the "insult" that was initially responsible for folate malabsorption.

The clinical pattern of tropical sprue varies in different areas of the world (e.g., India vs. Puerto Rico). Not infrequently, individuals in India initially will report the occurrence of an acute enteritis before the development of steatorrhea and malabsorption. In contrast, in Puerto Rico, a most insidious onset of symptoms and a more dramatic response to antibiotics is seen than in some other locations. Tropical sprue in different areas of the world may not be the same disease; there may be similar clinical entities but with different etiologies.

Diagnosis The diagnosis of tropical sprue is best made by the presence of an abnormal small-intestinal mucosal biopsy in an individual with chronic diarrhea and evidence of malabsorption who is either residing or has recently lived in a tropic country. The small-intestinal biopsy in tropical sprue does not have pathognomonic features but resembles, and can often be indistinguishable from, that seen in celiac sprue (Fig. 286-4). The biopsy in tropical sprue will have less villus architectural alteration and more mononuclear cell infiltrate in the lamina propria. In contrast to celiac sprue, the histopathologic features of tropical sprue are present with a similar degree of severity throughout the small intestine, and a gluten-free diet does not result in either clinical or histopathologic improvement in tropical sprue.

TREATMENT

Broad-spectrum antibiotics and folic acid are most often curative, especially if the patient leaves the tropical area and does not return. Tetracycline should be used for up to 6 months and may be associated with improvement within 1 to 2 weeks. Folic acid alone will induce a hematologic remission as well as improvement in appetite, weight gain, and some morphologic changes in small intestinal biopsy. Because of the presence of marked folate deficiency, folic acid is most often given together with antibiotics.

SHORT BOWEL SYNDROME

This is a descriptive term for the myriad clinical problems that often occur following resection of varying lengths of small intestine. The factors that determine both the type and degree of symptoms include: (1) the specific segment (jejunum vs. ileum) resected, (2) the length of the resected segment, (3) the integrity of the ileocecal valve, (4) whether any large intestine has also been removed, (5) residual disease in the remaining small and/or large intestine (e.g., Crohn's disease, mesenteric artery disease), and (6) the degree of adaptation in the remaining intestine. Short bowel syndrome can occur at any age from neonates through the elderly.

Three different situations in adults demand intestinal resections: (1) mesenteric vascular disease including both atherosclerosis, thrombotic phenomena, and vasculitidies; (2) primary mucosal and submucosal disease, e.g., Crohn's disease; and (3) operations without preexisting small intestinal disease, such as trauma and jejunoileal bypass for

obesity.

Following resection of the small intestine, the residual intestine undergoes adaptation of both structure and function that may last for up to 6 to 12 months. Adaptation requires the continued intake of dietary nutrients and calories to stimulate it via direct contact with ileal mucosa, the release of one or more intestinal hormones, and pancreatic and biliary secretions. Thus, enteral nutrition and calorie administration must be maintained, especially in the early postoperative period, even if an extensive intestinal resection requiring total parenteral nutrition (TPN) was required. The subsequent ability of such patients to absorb nutrients will not be known for several months until after adaptation is completed.

Multiple factors besides the absence of intestinal mucosa (required for both lipid and fluid and electrolyte absorption) contribute to the diarrhea and steatorrhea in these patients. Removal of the ileum and especially the ileocecal valve is often associated with more severe diarrhea than jejunal resection. Without part or all of the ileum, diarrhea can be caused by an increase in bile acids entering the colon, leading to their stimulation of colonic fluid and electrolyte secretion. Absence of the ileocecal valve is also associated with a decrease in intestinal transit time and bacterial overgrowth from the colon. Lactose intolerance as a result of the removal of lactase-containing mucosa as well as gastric hypersecretion will also contribute to the diarrhea.

In addition to diarrhea and/or steatorrhea, a range of nonintestinal symptoms are also observed in some patients. A significant increase in renal calcium oxalate calculi is observed in patients with a small-intestinal resection with an intact colon and is due to an increase in oxalate absorption by the large intestine, with subsequent hyperoxaluria. Since oxalate is high in relatively few foods (e.g., spinach, rhubarb, tea), dietary restrictions alone are not adequate treatment. Cholestyramine, an anion-binding resin, and calcium have proved useful in reducing the hyperoxaluria. Similarly, an increase in cholesterol gall stones is seen that is related to a decrease in the bile acid pool size, which results in the generation of cholesterol supersaturation in gall bladder bile. Gastric hypersecretion of acid occurs in many patients following large resections of the small intestine. The etiology is unclear but may be related to either reduced hormonal inhibition of acid secretion or increased gastrin levels due to reduced small-intestinal catabolism of circulating gastrin. The resulting gastric acid secretion may be an important factor contributing to the diarrhea and steatorrhea. A reduced pH in the duodenum can inactivate pancreatic lipase and/or precipitate duodenal bile acids, thereby increasing steatorrhea, and an increase in gastric secretion can create a volume overload relative to the reduced small-intestinal absorptive capacity. Inhibition of gastric acid secretion with either proton pump inhibitors or H_2receptor antagonists can help in reducing the diarrhea and steatorrhea.

TREATMENT

Treatment of short bowel syndrome depends on the severity of symptoms and whether the individual is able to maintain caloric and electrolyte balance with oral intake alone. Initial treatment includes judicious use of opiates to reduce stool output and to establish an effective diet. An initial diet should be low-fat, high-carbohydrate to minimize the diarrhea from fatty acid stimulation of colonic fluid secretion. BothMCT (see above), a

low-lactose diet, and various fiber-containing diets should also be tried. In the absence of an ileocecal valve, the possibility of bacterial overgrowth must be considered and treated. If gastric acid hypersecretion is contributing to the diarrhea and steatorrhea, a proton pump inhibitor may be helpful. Usually none of these therapeutic approaches will provide an instant solution but will reduce disabling diarrhea.

The patient's vitamin and mineral status must also be monitored, and replacement therapy initiated, if indicated. Fat-soluble vitamins, folate, cobalamin, calcium, iron, magnesium, and zinc are the most critical factors to monitor on a regular basis. If these approaches are not successful, homeTPNrepresents an established therapy that can be maintained for many years. Intestinal transplantation is beginning to become established as a possible approach for individuals with extensive intestinal resection who cannot be maintained without TPN.

BACTERIAL OVERGROWTH SYNDROME

Bacterial overgrowth syndrome comprises a group of disorders with diarrhea, steatorrhea, and macrocytic anemia whose common feature is the proliferation of colon-type bacteria within the small intestine. This bacterial proliferation is due to stasis caused by impaired peristalsis (i.e., *functional stasis*), changes in intestinal anatomy (i.e., *anatomic stasis*), or direct communication between the small and large intestine. These conditions have also been referred to as *stagnant bowel syndrome* or *blind loop syndrome*.

Pathogenesis The manifestations of bacterial overgrowth syndromes are a direct consequence of the presence of increased amounts of a colonic-type bacterial flora, such as *E. coli* or *Bacteroides*, in the small intestine. *Macrocytic anemia* is due to cobalamin, not folate, deficiency. Most bacteria require cobalamin for growth, and increasing concentrations of bacteria use up the relatively small amounts of dietary cobalamin. *Steatorrhea* is due to impaired micelle formation as a consequence of a reduced intraduodenal concentration of bile acids and the presence of unconjugated bile acids. Certain bacteria, e.g., *Bacteroides*, deconjugate conjugated bile acids to unconjugated bile acids. In the presence of bacterial overgrowth, unconjugated bile acids will be absorbed more rapidly than conjugated bile acids, and, as a result, the intraduodenal concentration of bile acids will be reduced. In addition, theCMC of unconjugated bile acids is higher than that of conjugated bile acids, resulting in a decrease in micelle formation. *Diarrhea* is due, at least in part, to the steatorrhea, when it is present. However, some patients manifest diarrhea *without* steatorrhea, and it is assumed that the colonic-type bacteria in these patients are producing one or more bacterial enterotoxins that are responsible for fluid secretion and diarrhea.

Etiology The etiology of these different disorders is bacterial proliferation in the small intestinal lumen secondary to either anatomic or functional stasis or to a communication between the relatively sterile small intestine and the colon with its high levels of aerobic and anaerobic bacteria. Several examples of anatomic stasis have been identified: (1) one or more diverticula (both duodenal and jejunal) (Figure 286-3C); (2) fistulas and strictures related to Crohn's disease (Figure 286-3D); (3) a proximal duodenal afferent loop following a subtotal gastrectomy and gastrojejunostomy; (4) a bypass of the intestine, e.g., jejunoileal bypass for obesity; and (5) dilatation at the site of a previous

intestinal anastomosis. The common feature of all of these anatomic derangements is the presence of a segment(s) of intestine that is out of continuity of propagated peristalsis, resulting in stasis and bacterial proliferation. Bacterial overgrowth syndromes can also occur in the absence of an anatomic blind loop when functional stasis is present. The best example of impaired peristalsis and bacterial overgrowth in the absence of a blind loop is scleroderma, where motility abnormalities exist in both the esophagus and small intestine (Chap. 313). Functional stasis and bacterial overgrowth can also occur in association with diabetes mellitus and in the small intestine when a direct connection exists between the small and large intestine, including an ileocolonic resection, or occasionally following an enterocolic anastomosis that permits entry of bacteria into the small intestine as a result of bypassing the ileocecal valve.

Diagnosis The diagnosis may be suspected from the combination of a low serum cobalamin level and an elevated serum folate level as enteric bacteria frequently produce folate compounds that will be absorbed in the duodenum. Ideally, the diagnosis of the bacterial overgrowth syndrome is the demonstration of increased levels of aerobic and/or anaerobic colonic-type bacteria in a jejunal aspirate obtained by intubation. This specialized test is rarely available, and bacterial overgrowth is best established by a Schilling test (Table 286-6), which should be abnormal following the administration of ^{58}Co-labeled cobalamin, with or without the administration of intrinsic factor. Following the administration of tetracycline for 5 days, the Schilling test will become normal, confirming the diagnosis of bacterial overgrowth.

TREATMENT

Primary treatment should be directed, if at all possible, to the surgical correction of an anatomic blind loop. In the absence of functional stasis, it is important to define the anatomic relationships responsible for stasis and bacterial overgrowth. For example, bacterial overgrowth secondary to strictures, one or more diverticula, or a proximal afferent loop can potentially be cured by surgical correction of the anatomic state. In contrast, the functional stasis of scleroderma or certain anatomic stasis states (e.g., multiple jejunal diverticula), cannot be corrected surgically, and these conditions should be treated with broad-spectrum antibiotics. Tetracycline used to be the initial treatment of choice but with increasing resistance, other antibiotics such as metronidazole, amoxicillin/clavulinic acid (Augmentin), and cephalosporin have been employed. The antibiotic should be given for approximately 3 weeks or until symptoms remit. Since the natural history of these conditions is chronic, antibiotics should not be given continuously, and symptoms usually remit within 2 to 3 weeks of initial antibiotic therapy. Therapy need not be repeated until symptoms recur. In the presence of frequent recurrences several treatment strategies exist, but the use of antibiotics for 1 week per month whether or not symptoms are present is often most effective.

Unfortunately, therapy for bacterial overgrowth syndrome is largely empirical, with an absence of clinical trials on which to base decisions regarding the antibiotic to be used, the duration of treatment, and/or the best approach for treating recurrences. Bacterial overgrowth may also occur as a component of another chronic disease, e.g., Crohn's disease, radiation enteritis, or short bowel syndrome. Treatment of the bacterial overgrowth in these settings will not cure the underlying problem but may be very important in ameliorating a subset of clinical problems that are related to bacterial

overgrowth.

WHIPPLE'S DISEASE

Whipple's disease is a chronic multisystem disease associated with diarrhea, steatorrhea, weight loss, arthralgia, and central nervous system and cardiac problems that is caused by the bacteria *Tropheryma whippelii*. Until the identification of *T. whippelii* by polymerase chain reaction during the past decade, the hallmark of Whipple's disease had been the presence ofPAS-positive macrophages in the small intestine and other organs with evidence of disease. Long before the establishment of *T. whippelii* as the causative agent of Whipple's disease, gram-positive bacilli had been identified both within and outside of macrophages.

Etiology Whipple's disease is caused by a small gram-positive bacillus, *T. whippelii*. The bacillus, an actinobacterium, has low virulence but high infectivity, and relatively minimal symptoms are observed compared to the extent of the bacilli in multiple tissues.

Clinical Presentation The onset of Whipple's disease is insidious and is characterized by diarrhea, steatorrhea, abdominal pain, weight loss, migratory large-joint arthropathy, and fever as well as ophthalmologic and central nervous system symptoms. The development of dementia is a relatively late symptom and is an extremely poor prognostic sign, especially in patients who relapse following the induction of a remission with antibiotics. For unexplained reasons, the disease occurs primarily in middle-aged (50-year-old) Caucasian men. The steatorrhea in these patients is generally believed secondary to both small-intestinal mucosal injury and lymphatic obstruction secondary to the increased number ofPAS-positive macrophages in the lamina propria of the small intestine.

Diagnosis The diagnosis of Whipple's disease is suggested by a multisystem disease in a 50-year-old Caucasian male with diarrhea and steatorrhea. Obtaining tissue biopsies from the small intestine and/or other organs that may be involved (e.g., liver, lymph nodes, heart, eyes, central nervous system, or synovial membranes), based on the patient's symptoms, is the primary approach to establish the diagnosis of Whipple's disease. The presence ofPAS-positive macrophages containing the characteristic small (0.25´ 1 to 2 um) bacilli is suggestive of this diagnosis. However, Whipple's disease can be confused with the PAS-positive macrophages containing *M. avian* complex, which may be a cause of diarrhea in AIDS. The presence of the *T. whippelii* bacillus outside of macrophages is a more important indicator of active disease than their presence within the macrophages. *T. whippelii* has now been successfully grown in culture.

TREATMENT

The treatment for Whipple's disease is prolonged use of antibiotics. At the present time the drug of choice is double-strength trimethoprim/sulfamethoxazole for approximately 1 year.PAS-positive macrophages can persist following successful treatment, and the presence of bacilli outside of macrophages is indicative of persistent infection or an early sign of recurrence. Recurrence of disease activity, especially with dementia, is an extremely poor prognostic sign and requires an antibiotic that crosses the blood-brain barrier. If trimethoprim/sulfamethoxazole is not tolerated, chloramphenicol is an

appropriate second choice.

PROTEIN-LOSING ENTEROPATHY

Protein-losing enteropathy is not a specific disease but rather describes a group of gastrointestinal and nongastrointestinal disorders with hypoproteinemia and edema in the absence of either proteinuria or defects in protein synthesis, e.g., chronic liver disease. These diseases are characterized by excess protein loss into the gastrointestinal tract. Normally, about 10% of the total protein catabolism occurs via the gastrointestinal tract. Evidence of increased protein loss into the gastrointestinal tract has been established in>65 different diseases, which can be classified into three primary groups: (1) mucosal ulceration such that the protein loss primarily represents exudation across damaged mucosa, e.g., ulcerative colitis, gastrointestinal carcinomas, peptic ulcer; (2) nonulcerated mucosa but with evidence of mucosal damage so that the protein loss represents loss across epithelia with altered permeability, e.g., celiac sprue and Menetrier's disease in the small intestine and stomach, respectively; (3) lymphatic dysfunction, either representing primary lymphatic disease or secondary to partial lymphatic obstruction that may occur as a result of enlarged lymph nodes or cardiac disease.

Diagnosis The diagnosis of protein-losing enteropathy is suggested by the presence of peripheral edema and low serum albumin and globulin levels in the absence of renal and hepatic disease. It is extremely rare for an individual with protein-losing enteropathy to have selective loss of *only* albumin or *only* globulins. Therefore, marked reduction of serum albumin with normal serum globulins should not initiate an evaluation for protein-losing enteropathy but should suggest the presence of renal and/or hepatic disease. Likewise, reduced serum globulins with normal serum albumin levels is more likely a result of reduced globulin synthesis rather than enhanced globulin loss into the intestine. Documentation of an increase in protein loss into the gastrointestinal tract has been established by the administration of one of several radiolabeled proteins and its quantitation in stool during a 24- or 48-h period. Unfortunately, none of these radiolabeled proteins are available for routine clinical use.a_1-Antitrypsin, a protein that represents approximately 4% of total serum proteins and is resistant to proteolysis, can be used to document enhanced rates of serum protein loss into the intestinal tract but cannot be used to assess gastric protein loss due to its degradation in an acid milieu.a_1-Antitrypsin clearance is measured by determining stool volume and both stool and plasmaa_1-antitrypsin concentrations. In addition to the loss of protein via abnormal and distended lymphatics, peripheral lymphocytes may also be lost via lymphatics, resulting in a relative lymphopenia. Thus, the presence of lymphopenia in a patient with hypoproteinemia supports the presence of increased loss of protein into the gastrointestinal tract.

Patients with increased protein loss into the gastrointestinal tract from lymphatic obstruction often have steatorrhea and diarrhea. The steatorrhea is a result of altered lymphatic flow as lipid-containing chylomicrons exit from intestinal epithelial cells via intestinal lymphatics (Table 286-4;Fig. 286-4). In the absence of mechanical or anatomic lymphatic obstruction, instrinsic intestinal lymphatic dysfunction, with or without lymphatic dysfunction in the peripheral extremities, has been named *intestinal lymphangiectasia*. Similarly, about 50% of individuals with intrinsic peripheral lymphatic

disease (Milroy's disease) will also have intestinal lymphangiectasia and hypoproteinemia. Other than steatorrhea and enhanced protein loss into the gastrointestinal tract, all other aspects of intestinal absorptive function are normal in intestinal lymphangiectasia.

Other Causes Patients who appear to have idiopathic protein-losing enteropathy without any evidence of gastrointestinal disease should be examined for cardiac disease and especially right-sided valvular disease and chronic pericarditis (Chap. 236). On occasion, hypoproteinemia can be the only presentation for these two types of heart disease. Menetrier's disease (also called *hypertrophic gastropathy*) is an uncommon entity that involves the body and fundus of the stomach and is characterized by large gastric folds, reduced gastric acid secretion, and, at times, enhanced protein loss into the stomach.

TREATMENT

As excess protein loss into the gastrointestinal tract is most often secondary to a specific disease, treatment should be directed primarily to the underlying disease process and not to the hypoproteinemia. For example, if significant hypoproteinemia with resulting peripheral edema is present secondary to either celiac sprue or ulceratic colitis, a gluten-free diet or mesalamine, respectively, would be the initial therapy. When enhanced protein loss is secondary to lymphatic obstruction, it is critical to establish the nature of this obstruction. Identification of mesenteric nodes or lymphoma may be possible by imaging studies. Similarly, it is important to exclude cardiac disease as a cause of protein-losing enteropathy either by echosonography or, on occasion, by a right-heart catheterization.

The increased protein loss that occurs in intestinal lymphangiectasia is a result of distended lymphatics associated with lipid malabsorption. Treatment of the hypoproteinemia is accomplished by a low-fat diet and the administration ofMCTs(Table 286-3), which do not exit from the intestinal epithelial cells via lymphatics but are delivered to the body via the portal vein.

SUMMARY

A pathophysiologic classification of the many conditions that can produce malabsorption is given in Table 286-9. A summary of the pathophysiology of the various clinical manifestations of malabsorption is given inTable 286-10.

(Bibliography omitted in Palm version)

Back to Table of Contents

287. INFLAMMATORY BOWEL DISEASE - *Sonia Friedman*, *Richard S. Blumberg*

Inflammatory bowel disease (IBD) is an idiopathic and chronic intestinal inflammation. Ulcerative colitis (UC) and Crohn's disease (CD) are the two major types of IBD.

EPIDEMIOLOGY

The incidence ofIBDvaries within different geographic areas. Northern countries, such as the United States, United Kingdom, Norway, and Sweden, have the highest rates. The incidence rates ofUC andCD in the United States are about 11 per 100,000 and 7 per 100,000, respectively (Table 287-1). Countries in southern Europe, South Africa, and Australia have lower incidence rates: 2 to 6.3 per 100,000 for UC, and 0.9 to 3.1 per 100,000 for CD. In Asia and South America, IBD is rare; incidence rates of UC and CD are 0.5 and 0.08 per 100,000, respectively. The highest mortality in IBD patients is during the first years of disease and in long duration disease due to the risk of colon cancer. In a Swedish population study, the standardized mortality ratios for CD and UC were 1.51 and 1.37, respectively.

The peak age of onset ofUC andCD is between 15 and 30 years. A second peak occurs between the ages of 60 and 80. The male to female ratio for UC is 1:1 and for CD is 1.1 to 1.8:1. A two- to fourfold increased frequency of UC and CD in Jewish populations has been described in the United States, Europe, and South Africa. Furthermore, disease frequency differs within the Jewish populations. The prevalence ofIBD in Ashkenazi Jews is about twice that of Israeli-born, Sephardic, or Oriental Jews. The prevalence decreases progressively in non-Jewish Caucasian, African-American, Hispanic, and Asian populations. Urban areas have a higher prevalence of IBD than rural areas and high socioeconomic classes have a higher prevalence than lower socioeconomic classes.

The effects of cigarette smoking are different inUC andCD. The risk of UC in smokers is 40% that of nonsmokers. Additionally, former smokers have a 1.7-fold increased risk for UC than people who have never smoked. In contrast, smoking is associated with a twofold increased risk of CD. Oral contraceptives are also linked to CD; the relative risk of CD for oral contraceptive users is about 1.9. Appendectomy appears to be protective against UC but further studies are needed.

IBDruns in families. If a patient has IBD, the lifetime risk that a first-degree relative will be affected is ~10%. If two parents have IBD, each child has a 36% chance of being affected. In twin studies, 67% of monozygotic twins are concordant forCD and 20% are concordant forUC, whereas 8% of dizygotic twins are concordant for CD and none are concordant for UC. There is also concordance for anatomic site and clinical type of CD within families.

Additional evidence for genetic predisposition toIBDcomes from its association with certain genetic syndromes.UC andCD are both associated with Turner's syndrome, and Hermansky-Pudlak syndrome is associated with a granulomatous colitis. Glycogen storage disease type 1b can present with Crohn's-like lesions of the large and small bowel. Other immunodeficiency disorders, such as hypogammaglobulinemia, selective IgA deficiency, and hereditary angioedema, also exhibit an increased association with

IBD.

ETIOLOGY AND PATHOGENESIS

AlthoughIBD has been described as a clinical entity for over 100 years, its etiology and pathogenesis have not been definitively elaborated. Various studies have led to a consensus hypothesis that in genetically predisposed individuals, both exogenous factors (e.g., infectious agents, normal lumenal flora) and host factors (e.g., intestinal epithelial cell barrier function, vascular supply, neuronal activity) together cause a chronic state of dysregulated mucosal immune function that is further modified by specific environmental factors (e.g., smoking). Although it is possible that the chronic activation of the mucosal immune system may represent an appropriate response to a chronic unidentified infectious agent, a search for such an agent has thus far been unrewarding. As such, IBD must currently be considered an inappropriate response to either the endogenous microbial flora within the intestine, with or without some component of autoimmunity. Importantly, the normal intestine contains a significant concentration of immune cells in a chronic state of so-called *physiologic inflammation*, in which the gut is poised for, but actively restrained from, full immunologic responses. During the course of infections in the normal host, full activation of the gut-associated lymphoid tissue occurs but is rapidly superceded by downregulation of the immune response and tissue repair. In IBD this process is not regulated normally.

GENETIC CONSIDERATIONS

IBDis a polygenic disorder that gives rise to multiple clinical subgroups withinUC andCD. Genome-wide searches have shown that potential disease-associated loci are present on chromosomes 16, 12, 7, 3, and 1, although the specific gene associations are undefined. HLA alleles may play a role. UC patients disproportionately express DR2-related alleles, whereas in CD an increased use of the DR5 DQ1 haplotype or the DRB*0301 allele has been described. In UC patients with pancolitis undergoing total proctocolectomy, 14.3% versus 3.2% of non-IBD controls express the HLA DRB1*0103 allele. This allele is associated with extensive disease and extraintestinal manifestations such as mouth ulcers, arthritis, and uveitis. Other associations with immunoregulatory genes include the intercellular adhesion molecule R241 allele in UC and CD and the interleukin (IL) 1 receptor antagonist allele 2 in UC patients that is associated with total colonic inflammation. Although not proven at the genetic level, patients with IBD and their first-degree relatives may exhibit diminished intestinal epithelial cell barrier function.

DEFECTIVE IMMUNE REGULATION IN IBD

The normal state of the mucosal immune system is one of inhibited immune responses to lumenal contents due to oral tolerance that occurs in the normal individual. When soluble antigens are administered orally rather than subcutaneously or intramuscularly, antigen-specific non-responsiveness is induced. Multiple mechanisms are involved in the induction of oral tolerance and include deletion or anergy of antigen-reactive T cells or activation of CD4+ T cells that suppress gut inflammation through secretion of inhibitory cytokines (IL-10, TGF-b). Oral tolerance may be responsible for the lack of immune responsiveness to dietary antigens and the commensal flora in the intestinal

lumen. In IBD this tightly regulated state of suppression of inflammation is altered, leading to uncontrolled inflammation. The mechanisms that maintain this regulated state of immune suppression are unknown.

Gene knockout (-/-) or transgenic (Tg) mouse models of colitis have revealed that deleting specific cytokines (e.g.,IL-2, IL-10, TGF-b) or their receptors, deleting molecules associated with T-cell antigen recognition (e.g., T-cell antigen receptors, MHC class II), or interfering with intestinal epithelial cell barrier function (e.g., blocking N-cadherin, deleting multidrug resistance gene 1a or trefoil factor) leads to colitis. Thus, a variety of specific alterations can lead to unregulated autoimmunity directed at the colon in mice.

In bothUC andCD, activated CD4+ T cells present in the lamina propria and peripheral blood secrete inflammatory cytokines. Some directly activate other inflammatory cells (macrophages and B cells) and others act indirectly to recruit other lymphocytes, inflammatory leukocytes, and mononuclear cells from the peripheral vasculature into the gut through interactions between homing receptors on leukocytes (e.g.,a4b7 integrin) and addressins on vascular endothelium (e.g., MadCAM1). CD4+ T cells can be subdivided into two major categories both of which may be associated with colitis in animal models and humans: T_H1 cells (IFN-g, TNF) and T_H2 cells (IL-4, IL-5, IL-13). T_H1 cells appear to induce transmural granulomatous inflammation that resembles CD, and T_H2 cells appear to induce superficial mucosal inflammation more characteristic of UC. The T_H1 cytokine pathway is initiated by IL-12, a key cytokine in the pathogenesis of experimental models of mucosal inflammation. Thus, use of antibodies to block proinflammatory cytokines (e.g., anti-TNF-a, anti-IL-12) or molecules associated with leukocyte recruitment (e.g., anti-a4b7) or use of cytokines that inhibit inflammation (e.g., IL-10) or promote intestinal barrier function (e.g., IL-11) may be beneficial to humans with colitis.

THE INFLAMMATORY CASCADE INIBD

Once initiated in IBD, the immune inflammatory response is perpetuated as a consequence of T-cell activation. A sequential cascade of inflammatory mediators acts to extend the response; each step is a potential target for therapy. Inflammatory cytokines, such asIL-1, IL-6, and tumor necrosis factor (TNF) have diverse effects on tissue. They promote fibrogenesis, collagen production, activation of tissue metalloproteinases, and the production of other inflammatory mediators; they also activate the coagulation cascade in local blood vessels (e.g., increased production of von Willebrand's factor). These cytokines are normally produced in response to infection, but are usually turned off or inhibited at the appropriate time to limit tissue damage. In IBD their activity is not regulated, resulting in an imbalance between the proinflammatory and anti-inflammatory mediators. Therapies such as the 5-ASA compounds are potent inhibitors of these inflammatory mediators through inhibition of transcription factors such as NF-kB that regulate their expression.

EXOGENOUS FACTORS

IBDmay have an as yet undefined infectious etiology. Three specific agents have received the greatest attention, *Mycobacterium paratuberculosis*, *Paramyxovirus*, and

Helicobacter species. The immune response to a specific organism could be expressed differently, depending upon the individual's genetic background. Although *M. paratuberculosis* had initially been identified inCDpatients, further studies have not confirmed a disease association. In addition, antimycobacterial agents have not been effective in treating CD. A role for the measles virus or paramyxoviruses in the development of CD has been suggested based on an increase in the incidence of CD in England that paralleled use of the measles vaccine. However, studies in the United States have not substantiated this finding. In an animal model of IBD, *Helicobacter hepaticus* has been implicated as a trigger for the inflammatory response; evidence in people is lacking.

Multiple pathogens (e.g., *Salmonella*, *Shigella sp.*, *Campylobacter sp.*) may initiateIBD by triggering an inflammatory response that the mucosal immune system may fail to control. However, in an IBD patient the normal flora is likely perceived as if it were a pathogen. Anaerobic organisms, particularly *Bacteroides* species, may be responsible for the induction of inflammation. Such a notion is supported by the response in patients withCD to agents that alter the intestinal flora, such as metronidazole, ciprofloxacin, and elemental diets. CD also responds to fecal diversion, demonstrating the ability of lumenal contents to exacerbate disease. On the other hand, other bacterial organisms, so-called probiotics such as *Lactobacillus sp.*, downregulate inflammation in animal models and humans.

Psychosocial factors can contribute to clinical exacerbation of symptoms. Major life events such as illness or death in the family, divorce or separation, interpersonal conflict, or other major loss, are associated with an increase inIBDsymptoms such as pain, bowel dysfunction, and bleeding. Acute daily stress can exacerbate bowel symptoms even after controlling for major life events. When the *sickness impact profile*, a measurement of overall psychological and physical functioning is used, IBD patients have functional impairment greater than that of a health maintenance organization population but less than that of patients with chronic back pain or amyotrophic lateral sclerosis. IBD patients have been hypothesized to have a characteristic personality that renders them susceptible to emotional stresses. However, emotional dysfunction could also be the result of chronic illness and should be considered when treating these patients.

PATHOLOGY

ULCERATIVE COLITIS: MACROSCOPIC FEATURES

UCis a mucosal disease that usually involves the rectum and extends proximally to involve all or part of the colon. Approximately 40 to 50% of patients have disease limited to the rectum and rectosigmoid, 30 to 40% have disease extending beyond the sigmoid but not involving the whole colon, and 20% have a total colitis. Proximal spread occurs in continuity without areas of uninvolved mucosa. When the whole colon is involved, the inflammation extends 1 to 2 cm into the terminal ileum in 10 to 20% of patients. This is called *backwash ileitis* and has little clinical significance. Although variations in macroscopic activity may suggest skip areas, biopsies from normal-appearing mucosa are usually abnormal. Thus, it is important to obtain multiple biopsies from apparently uninvolved mucosa, whether proximal or distal, during endoscopy.

With mild inflammation, the mucosa is erythematous and has a fine granular surface that looks like sandpaper. In more severe disease, the mucosa is hemorrhagic, edematous, and ulcerated (Fig. 287-1). In long-standing disease, inflammatory polyps (pseudopolyps) may be present as a result of epithelial regeneration (Fig. 287-2). The mucosa may appear normal in remission but in patients with many years of disease it appears atrophic and featureless and the entire colon becomes narrowed and foreshortened (Fig. 287-3). Patients with fulminant disease can develop a toxic colitis or a toxic megacolon where the bowel wall becomes very thin and the mucosa is severely ulcerated, which may lead to perforation.

ULCERATIVE COLITIS: MICROSCOPIC FEATURES

Histologic findings correlate well with the endoscopic appearance and clinical course ofUC. The process is limited to the mucosa and superficial submucosa with deeper layers unaffected except in fulminant disease (Fig. 287-4). In UC, two major histologic features are indicative of chronicity and help distinguish it from infectious or acute self-limited colitis. First, the crypt architecture of the colon is distorted; crypts may be bifid and reduced in number, often with a gap between the crypt bases and the muscularis mucosae. Second, some patients have basal plasma cells and multiple basal lymphoid aggregates. Mucosal vascular congestion with edema and focal hemorrhage, and an inflammatory cell infiltrate of neutrophils, lymphocytes, plasma cells, and macrophages may be present. The neutrophils invade the epithelium, usually in the crypts, and give rise to cryptitis and, ultimately, to crypt abscesses (Fig. 287-5). The cryptitis is associated with mucus discharge from goblet cells and increased epithelial cell turnover. Histologically, this results in goblet cell depletion. Other chronic changes that are sometimes seen are neuronal hypertrophy and fibromuscular hyperplasia of the muscularis mucosae.

CROHN'S DISEASE: MACROSCOPIC FEATURES

CDcan affect any part of the gastrointestinal tract from the mouth to the anus. Some 30 to 40% of patients have small bowel disease alone, 40 to 55% have disease involving both the small and large intestines, and 15 to 25% have colitis alone. In the 75% of patients with small intestinal disease, the terminal ileum is involved in 90%. UnlikeUC, which almost always involves the rectum, the rectum is often spared in CD. CD is segmental, with skip areas in the midst of diseased intestine (Fig. 287-6). Perirectal fistulas, fissures, abscesses, and anal stenosis are present in one-third of patients with CD, particularly those with colonic involvement. CD may also involve the liver and the pancreas.

UnlikeUC,CD is a transmural process (Fig. 287-7). Endoscopically, aphthous or small superficial ulcerations characterize mild disease; in more active disease, stellate ulcerations fuse longitudinally and transversely to demarcate islands of mucosa that frequently are histologically normal. This "cobblestone" appearance is characteristic of CD, both endoscopically and by barium radiography. As in UC, pseudopolyps can form in CD.

Active CD is characterized by focal inflammation and formation of fistula tracts, which

resolve by fibrosis and stricturing of the bowel. The bowel wall thickens and becomes narrowed and fibrotic, leading to chronic, recurrent bowel obstructions. Projections of thickened mesentery encase the bowel ("creeping fat") and serosal and mesenteric inflammation promote adhesions and fistula formation.

CROHN'S DISEASE: MICROSCOPIC FEATURES

The earliest lesions are aphthoid ulcerations and focal crypt abscesses with loose aggregations of macrophages, which form noncaseating granulomas in all layers of the bowel wall from mucosa to serosa (Fig. 287-8). Granulomas can be seen in lymph nodes, mesentery, peritoneum, liver, and pancreas. Although granulomas are a pathognomonic feature ofCD, only half of cases reveal granulomas on surgical or endoscopic biopsy specimens. Other histologic features of CD include submucosal or subserosal lymphoid aggregates, particularly away from areas of ulceration, gross and microscopic skip areas, and transmural inflammation that is accompanied by fissures that penetrate deeply into the bowel wall and sometimes form fistulous tracts or local abscesses.

CLINICAL PRESENTATION

ULCERATIVE COLITIS

Signs and Symptoms The major symptoms ofUC are diarrhea, rectal bleeding, tenesmus, passage of mucus, and crampy abdominal pain. The severity of symptoms correlates with the extent of disease. Although UC can present acutely, symptoms usually have been present for weeks to months. Occasionally, diarrhea and bleeding are so intermittent and mild that the patient does not seek medical attention.

Patients with proctitis usually pass fresh blood or blood-stained mucus, either mixed with stool or streaked onto the surface of a normal or hard stool. They also have tenesmus, or urgency with a feeling of incomplete evacuation. They rarely have abdominal pain. With proctitis or proctosigmoiditis, proximal transit slows, which may account for the constipation that is commonly seen in patients with distal disease.

When the disease extends beyond the rectum, blood is usually mixed with stool, or grossly bloody diarrhea may be noted. Colonic motility is altered by inflammation with rapid transit through the inflamed intestine. When the disease is severe, patients pass a liquid stool containing blood, pus, and fecal matter. Diarrhea is often nocturnal and/or postprandial. Although severe pain is not a prominent symptom, some patients with active disease may experience vague lower abdominal discomfort or mild central abdominal cramping. Severe cramping and abdominal pain can occur in association with severe attacks of the disease. Other symptoms in moderate to severe disease include anorexia, nausea, vomiting, fever, and weight loss.

Physical signs of proctitis include a tender anal canal and blood on rectal exam. With more extensive disease, patients have tenderness to palpation directly over the colon. Patients with a toxic colitis have severe pain and bleeding, and those with megacolon have hepatic tympany. Both may have signs of peritonitis if a perforation has occurred. The classification of disease activity is shown inTable 287-2.

Laboratory, Endoscopic, and Radiographic Features Active disease can be associated with a rise in acute phase reactants (C-reactive protein, orosomucoid levels), platelet count, erythrocyte sedimentation rate (ESR) and a decrease in hemoglobin. In severely ill patients, the serum albumin level will fall rather quickly. Leukocytosis may be present but is not a specific indicator of disease activity. Proctitis or proctosigmoiditis rarely causes a rise in C-reactive protein. Diagnosis relies upon the patient's history; clinical symptoms, negative stool examination for bacteria, *Clostridium difficile* toxin, and ova and parasites; sigmoidoscopic appearance; and histology of rectal or colonic biopsy specimens.

Sigmoidoscopy is used to assess disease activity and is often performed before treatment. Histologic features change more slowly than clinical features but can also be used to grade disease activity (Table 287-3).

Patients with a severe attack ofUCshould have a plain, supine film of the abdomen. In the presence of severe disease, the margin of the colon becomes edematous and irregular (Fig. 287-9). Colonic thickening and toxic dilation can both be seen on a plain radiograph.

The earliest radiologic change ofUC seen on single-contrast barium enema is a fine mucosal granularity (Fig. 287-10). With increasing severity, the mucosa becomes thickened and superficial ulcers are seen. Deep ulcerations can appear as "collar-button" ulcers, which indicate that the ulceration has penetrated the mucosa. Haustral folds may be normal in mild disease, but as activity progresses they become edematous and thickened. Loss of haustration can occur, especially in patients with long-standing disease. In addition, the colon becomes shortened and narrowed. Polyps in the colon may be postinflammatory polyps or pseudopolyps (Fig. 287-11), adenomatous polyps, or carcinoma.

Computed tomography (CT) scanning is not as helpful as endoscopy and barium enema in making the diagnosis ofUC, but typical findings include mild mural thickening (<1.5 cm), inhomogeneous wall density, absence of small bowel thickening, increased perirectal and presacral fat, target appearance of the rectum, and adenopathy.

Complications Only 15% of patients withUCpresent initially with catastrophic illness. Massive hemorrhage occurs with severe attacks of disease in 1% of patients and treatment for the disease usually stops the bleeding. However, if patients require 6 to 8 units of blood within 24 to 48 h, colectomy is indicated. Toxic megacolon is defined as a transverse colon with a diameter of more than 5.0 cm to 6.0 cm, with loss of haustration in patients with severe attacks of UC. It occurs in about 5% of attacks and can be triggered by electrolyte abnormalities and narcotics. Approximately 50% of acute dilations will resolve with medical therapy alone, but urgent colectomy is required for those that do not improve. Perforation is the most dangerous of the local complications, and the physical signs of peritonitis may not be obvious, especially if the patient is receiving glucocorticoids. Although perforation is rare, the mortality rate for perforation complicating a toxic megacolon is about 15%. In addition, patients can develop a toxic colitis and such severe ulcerations that the bowel may perforate without first dilating.

Obstructions caused by benign stricture formation occur in 10% of patients, with one-third of the strictures occurring in the rectum. These should be surveyed endoscopically for carcinoma.UCpatients occasionally develop anal fissures, perianal abscesses, or hemorrhoids but the occurrence of extensive perianal lesions should suggestCD.

CROHN'S DISEASE

Signs and Symptoms AlthoughCDusually presents as acute or chronic bowel inflammation, the inflammatory process evolves toward one of two patterns of disease: a fibrostenotic-obstructing pattern or a penetrating-fistulous pattern, each with different treatments and prognoses. The site of disease influences the clinical manifestations.

Ileocolitis Because the most common site of inflammation is the terminal ileum, the usual presentation of ileocolitis is a chronic history of recurrent episodes of right lower quadrant pain and diarrhea. Sometimes the initial presentation mimics acute appendicitis with pronounced right lower quadrant pain, a palpable mass, fever, and leukocytosis. Only at laparotomy, when the appendix is found to be normal, is the ileitis discovered. Pain is usually colicky; it precedes and is relieved by defecation. A low-grade fever is usually noted. High-spiking fever suggests intraabdominal abscess formation. Weight loss is common -- typically 10 to 20% of body weight -- and develops as a consequence of diarrhea, anorexia, and fear of eating.

An inflammatory mass may be palpated in the right lower quadrant of the abdomen. The mass is composed of inflamed bowel, adherent and indurated mesentery, and enlarged abdominal lymph nodes. Extension of the mass can cause obstruction of the right ureter or bladder inflammation, manifested by dysuria and fever. Edema, bowel wall thickening, and fibrosis of the bowel wall within the mass account for the radiographic "string sign" of a narrowed intestinal lumen.

Bowel obstruction may take several forms. In the early stages of the disease, bowel wall edema and spasm produce intermittent obstructive manifestations and increasing symptoms of postprandial pain. Over several years, this persistent inflammation gradually progresses to fibrostenotic narrowing and stricture. Diarrhea will decrease and eventually lead to chronic bowel obstruction and obstipation. Acute episodes of obstruction occur as well, precipitated by bowel inflammation and spasm or sometimes by impaction of undigested food. These episodes usually resolve with intravenous fluids and gastric decompression.

Severe inflammation of the ileocecal region may lead to localized wall thinning, with microperforation and fistula formation to the adjacent bowel, the skin, the urinary bladder, or to an abscess cavity in the mesentery. Enterovesical fistulas typically present as dysuria or recurrent bladder infections or less commonly as pneumaturia or fecaluria. Enterocutaneous fistulas follow tissue planes of least resistance, usually draining through abdominal surgical scars. Enterovaginal fistulas are rare and present as dyspareunia or as a feculent or foul-smelling, often painful vaginal discharge. They are unlikely to develop without a prior hysterectomy.

Jejunoileitis Extensive inflammatory disease is associated with a loss of digestive and

absorptive surface, resulting in malabsorption and steatorrhea. Nutritional deficiencies can also result from poor intake and enteric losses and protein and other nutrients. Intestinal malabsorption can cause hypoalbuminemia, hypocalcemia, hypomagnesemia, coagulopathy, and hyperoxaluria with nepholithiasis. Vertebral fractures are caused by a combination of vitamin D deficiency, hypocalcemia, and prolonged glucocorticoid use. Pellagra from niacin deficiency has been reported in extensive small bowel disease, and malabsorption of vitamin B12 can lead to a megaloblastic anemia.

Diarrhea is characteristic of active disease; its causes include: (1) bacterial overgrowth in obstructive stasis or fistulization, (2) bile-acid malabsorption due to a diseased or resected terminal ileum, (3) intestinal inflammation with decreased water absorption and increased secretion of electrolytes.

Colitis and Perianal Disease Patients with colitis present with low-grade fevers, malaise, diarrhea, crampy abdominal pain, and sometimes hematochezia. Gross bleeding due to deep colonic ulceration is not as common as inUC and appears in about half of patients with exclusively colonic disease. Only 1 to 2% bleed massively. Pain is caused by passage of fecal material through narrowed and inflamed segments of large bowel. Decreased rectal compliance is another cause for diarrhea in Crohn's colitis patients. Toxic megacolon has been associated with severe inflammation and short-duration disease.

Stricturing can occur in the colon, and patients can develop fibrous strictures with symptoms of bowel obstruction. Also, colonic disease may fistulize into the stomach or duodenum, causing feculent vomiting, or to the proximal or mid small bowel, causing malabsorption by "short circuiting" and bacterial overgrowth. Approximately 10% of women with Crohn's colitis will develop a rectovaginal fistula.

Perianal disease affects about one-third of patients with Crohn's colitis and is manifested by incontinence, large hemorrhoidal tags, anal strictures, anorectal fistulae, and perirectal abscesses. Not all patients with perianal fistula will have endoscopic evidence of colonic inflammation.

Gastroduodenal Disease Symptoms and signs of upper gastrointestinal tract disease include nausea, vomiting, and epigastric pain. Patients usually have a *H. pylori*-negative gastritis. The second portion of the duodenum is more commonly involved than the bulb. Fistulas involving the stomach or duodenum arise from the small or large bowel and do not necessarily signify the presence of upper gastrointestinal tract involvement. Patients with advanced gastroduodenalCD may develop a chronic gastric outlet obstruction.

Laboratory, Endoscopic, and Radiographic Features Laboratory abnormalities include elevated sedimentation rate and C-reactive protein. In more severe disease, findings include hypoalbuminemia, anemia, and leukocytosis.

Endoscopic features ofCDinclude rectal sparing, aphthous ulcerations, fistulas, and skip lesions. Endoscopy is useful for biopsy of mass lesions or strictures, or for visualization of filling defects seen on barium enema. Colonoscopy allows examination and biopsy of the terminal ileum, and upper endoscopy is useful in diagnosing gastroduodenal involvement in patients with upper tract symptoms. Ileal or colonic strictures may be

dilated with balloons introduced through the colonoscope. Endoscopic appearance correlates poorly with clinical remission; thus, repeated endoscopy is not used to monitor the inflammation.

InCD early radiographic findings in the small bowel include thickened folds and aphthous ulcerations. "Cobblestoning" from longitudinal and transverse ulcerations most frequently involves the small bowel (Fig. 287-12). In more advanced disease, strictures, fistulas (Fig. 287-13), inflammatory masses, and abscesses may be detected. The earliest macroscopic findings of colonic CD are aphthous ulcers. These small ulcers are often multiple and separated by normal intervening mucosa. As more severe disease develops, aphthous ulcers become enlarged, deeper, and occasionally connected to one another, forming longitudinal stellate, serpiginous, and linear ulcers.

The transmural inflammation ofCD leads to decreased luminal diameter and limited distensibility. As ulcers progress deeper, they can lead to fistula formation. The radiographic "string sign" (Figs. 287-14 and287-15) represents long areas of circumferential inflammation and fibrosis, resulting in long segments of luminal narrowing. The segmental nature of CD results in wide gaps of normal or dilated bowel between involved segments (Fig. 287-16).

CTfindings include mural thickening >2 cm, homogeneous wall density, mural thickening of small bowel, mesenteric fat stranding, perianal disease, and adenopathy. CT scanning can help identify abscesses, fistulas, and sinus tracts. Magnetic resonance imaging (MRI) may prove superior for demonstrating pelvic lesions such as ischiorectal abscesses.

Complications BecauseCD is a transmural process, serosal adhesions develop that provide direct pathways for fistula formation and reduce the incidence of free perforation. Free perforation occurs in 1 to 2% of patients, usually in the ileum but occasionally in the jejunum or as a complication of toxic megacolon. The peritonitis of free perforation, especially colonic, may be fatal. Generalized peritonitis may also result from the rupture of an intraabdominal abscess. Other complications include intestinal obstruction in 40%, massive hemorrhage, malabsorption, and severe perianal disease.

Serologic Markers Several serologic markers may be used to differentiate betweenCD andUC and help to predict the course of disease. Two antibodies that can be detected in the serum ofIBDpatients are perinuclear antineutrophil cytoplasmic antibody (pANCA) and anti-*Saccharomyces cerevisiae* antibodies (ASCA). A distinct set of antineutrophil cytoplasmic antibodies with perinuclear staining by indirect immunofluorescence is associated with UC. The antigens to which these antibodies are directed have not been identified, but they are distinct from those associated with vasculitis and may be related to histones. pANCA positivity is found in about 60 to 70% of UC patients and 5 to 10% of CD patients; 5 to 15% of first-degree relatives of UC patients are pANCA positive, whereas only 2 to 3% of the general population is pANCA positive. pANCA may also identify specific disease phenotypes. pANCA positivity is more often associated with pancolitis, early surgery, pouchitis, or inflammation of the pouch after ileal pouch-anal anastamosis (IPAA) and primary sclerosing cholangitis. pANCA in CD is associated with colonic disease that resembles UC.

ASCAantibodies recognize mannose sequences in the cell wall mannan of *S. cerevisiae*; 60 to 70% ofCDpatients, 10 to 15% of UC patients, and up to 5% of non-IBDcontrols are ASCA positive. The combined measurement ofpANCAand ASCA has been advocated as a valuable diagnostic approach to IBD. In one report, pANCA positivity with ASCA negativity yielded a 57% sensitivity and 97% specificity for UC, whereas pANCA negativity with ASCA positivity yielded a 49% sensitivity and 97% specificity for CD. ASCA was associated with small bowel CD. These antibody tests may help decide whether a patient with indeterminate colitis should undergo an IPAA, because patients with predominant features of CD often have a more difficult postoperative course.

Anti-goblet cell autoantibodies (GABs) -- autoantibodies against two target antigens in colonic epithelial cells -- are present in 39% ofUCpatients, 30% ofCDpatients, 21% of first-degree relatives of UC patients, 19% of first-degree relatives of CD patients, and 2% of healthy controls. An anti-colon antibody is found in 36% of UC patients and 13% of CD patients and healthy controls. In addition, 31% of CD patients and 4% of UC patients have serum antibodies against pancreatic acinar cells or pancreatic autoantibodies (PABs). Antibodies to red cell membrane antigens that cross-react with enteropathogens such as *Campylobacter sp.* may be associated with hemolytic anemia in CD. None of these antibodies are useful in the diagnosis and management of patients withIBD.

DIFFERENTIAL DIAGNOSIS OF UC AND CD

UC and CD have similar features to many other diseases. In the absence of a key diagnostic test, a combination of clinical, laboratory, histopathologic, radiographic, and therapeutic observations is required (Table 287-4). Once a diagnosis ofIBD is made, distinguishing between UC and CD is impossible in 10 to 20% of cases. These are termed *indeterminate* colitis.

INFECTIOUS DISEASE

Infections of the small intestines and colon can mimicCD orUC. They may be bacterial, fungal, viral, or protozoal in origin (Table 287-5). *Campylobacter* colitis can mimic the endoscopic appearance of severe UC and can cause a relapse of established UC. *Salmonella* can cause watery or bloody diarrhea, nausea, and vomiting. Shigellosis causes watery diarrhea, abdominal pain, and fever followed by rectal tenesmus and by the passage of blood and mucus per rectum. All three are usually self-limited but 1% of patients infected with *Salmonella* become asymptomatic carriers. *Yersinia enterocolitica* infection occurs mainly in the terminal ileum and causes mucosal ulceration, neutrophil invasion, and thickening of the ileal wall. Other bacterial infections that may mimicIBDinclude *C. difficile*, which presents with watery diarrhea, tenesmus, nausea, and vomiting, and *Escherichia coli*, three categories of which can cause colitis. These are enterohemorrhagic, enteroinvasive, and enteroadherent *E. coli*, all of which can cause bloody diarrhea and abdominal tenderness. Diagnosis of bacterial colitis is made by sending stool specimens for bacterial culture and *C. difficile* toxin analysis. Gonorrhea, *Chlamydia*, and syphilis can also cause proctitis

Gastrointestinal involvement with mycobacterial infection occurs primarily in the

immunosuppressed patient but may occur in patients with normal immunity. Distal ileal and cecal involvement predominates and patients present with symptoms of small bowel obstruction and a tender abdominal mass. The diagnosis is made most directly by colonoscopy with biopsy and culture. *Mycobacterium avium intracellulare* complex infection occurs in advanced stages of HIV infection and in other profoundly immunocompromised states, and usually manifests as a systemic infection with diarrhea, abdominal pain, weight loss, fever, and malabsorption. Diagnosis is established by acid-fast smear and culture of mucosal biopsies.

Although most of the patients with viral colitis are immunosuppressed, cytomegalovirus (CMV) and herpes simplex proctitis may occur in immunocompetent individuals. CMV occurs most commonly in the esophagus, colon, and rectum, but may also involve the small intestine. Symptoms include abdominal pain, bloody diarrhea, fever, and weight loss. With severe disease, necrosis and perforation can occur. Diagnosis is made by identification of intranuclear inclusions in mucosal cells on biopsy. Herpes simplex infection of the gastrointestinal tract is limited to the oropharynx, anorectum, and perianal areas. Symptoms include anorectal pain, tenesmus, constipation, inguinal adenopathy, difficulty with urinary voiding, and sacral paresthesias. Diagnosis is made by rectal biopsy. HIV itself can cause diarrhea, nausea, vomiting, and anorexia. Small intestinal biopsies show partial villus atrophy; small bowel bacterial overgrowth and fat malabsorption may also be noted.

Protozoan parasites include *Isospora belli*, which can cause a self-limited infection in healthy hosts but causes a chronic profuse, watery diarrhea and weight loss in AIDS patients. *Entamoeba histolytica* or related species infect about 10% of the world's population; symptoms include abdominal pain, tenesmus, frequent loose stool containing blood and mucus, and abdominal tenderness. Colonoscopy reveals focal punctate ulcers with normal intervening mucosa; diagnosis is made by biopsy or serum amebic antibodies. Fulminant amebic colitis is rare but has a mortality rate of>50%.

Other parasitic infections that may mimic IBD include hookworm (*Necator americanus*), whipworm (*Trichuris trichiura*), and *Strongyloides stercoralis*. In severely immunocompromised patients *Candida* or *Aspergillus* can be identified in the submucosa. Disseminated histoplasmosis can involve the ileocecal area.

NONINFECTIOUS DISEASE

Many diseases may mimicIBD(Table 287-5). Diverticulitis can be confused withCDclinically and radiographically. Both diseases cause fever, abdominal pain, tender abdominal mass, leukocytosis, elevatedESR, partial obstruction, and fistulas. Perianal disease or ileitis on small bowel series favors the diagnosis of CD. Significant endoscopic mucosal abnormalities are more likely in CD than in diverticulitis. Endoscopic or clinical recurrence following segmental resection favors CD. Diverticular-associated colitis is similar to CD, but mucosal abnormalities are limited to the sigmoid and descending colon.

Ischemic colitis is commonly confused withIBD. The ischemic process can be chronic and diffuse as inUC, or segmental as inCD. Colonic inflammation due to ischemia may resolve quickly or may persist and result in transmural scarring and stricture formation.

Ischemic bowel disease should be considered in the elderly following abdominal aortic aneurysm repair or when a patient has a hypercoagulable state or a severe cardiac or peripheral vascular disorder. Patients usually present with sudden onset of left lower quadrant pain, urgency to defecate, and the passage of bright red blood per rectum. Endoscopic examination often demonstrates a normal-appearing rectum and a sharp transition to an area of inflammation in the descending colon and splenic flexure.

The effects of radiation therapy on the gastrointestinal tract can be difficult to distinguish fromIBD. Acute symptoms can occur within 1 to 2 weeks of starting radiotherapy. When the rectum and sigmoid are irradiated, patients develop bloody, mucoid diarrhea and tenesmus, as in distalUC. With small bowel involvement, diarrhea is common. Late symptoms include malabsorption and weight loss. Stricturing with obstruction and bacterial overgrowth may occur. Fistulas can penetrate the bladder, vagina, or abdominal wall. Flexible sigmoidoscopy reveals mucosal granularity, friability, numerous telangiectasias, and occasionally discrete ulcerations. Biopsy can be diagnostic.

Solitary rectal ulcer syndrome is uncommon and can be confused withIBD. It occurs in mostly young females and may be caused by impaired evacuation and failure of relaxation of the puborectalis muscle. Ulceration may arise from anal sphincter overactivity, higher intrarectal pressures during defecation, and digital removal of stool. Patients complain of constipation with straining and pass blood and mucus per rectum. Other symptoms include abdominal pain, diarrhea, tenesmus, and perineal pain. The ulceration, which can be as large as 5 cm in diameter, is usually seen anteriorly or anteriorlaterally 3 to 15 cm from the anal verge. Biopsies can be diagnostic.

Several types of colitis have been associated with nonsteroidal anti-inflammatory drugs (NSAID), including de novo colitis, reactivation ofIBD, and proctitis caused by use of suppositories. Most patients with NSAID-related colitis present with diarrhea and abdominal pain and complications include stricture, bleeding, obstruction, perforation, and fistulization. Withdrawal of these agents is crucial, and in cases of reactivated IBD, standard therapies are indicated.

INDETERMINITE COLITIS

Cases ofIBD that cannot be categorized asUC orCD are called *indeterminate* colitis. Long-term follow-up reduces the number of patients labeled indeterminate to about 10%. The disease course of indeterminate colitis is unclear and surgical recommendations are difficult, especially since up to 20% of pouches fail, requiring ileosotomy. A multistage ileal pouch-anal anastamosis (the initial stage consisting of a subtotal colectomy with Hartman pouch) with careful histologic evaluation of the resected specimen to exclude CD is advised. Medical therapy is similar to UC and CD; most clinicians use 5-ASA drugs, glucocorticoids, and immunomodulators as necessary.

THE ATYPICAL COLITIDIES

Two atypical colitides -- collagenous colitis and lymphocytic colitis -- have completely normal endoscopic appearances. Collagenous colitis has two main histologic components: increased subepithelial collagen deposition and colitis with increased intraepithelial lymphocytes. Female to male ratio is 9:1, and most patients present in the

sixth or seventh decades of life. The main symptom is chronic watery diarrhea. Treatments range from sulfasalazine and lomotil to bismuth to glucocorticoids for refractory disease.

Lymphocytic colitis has features similar to collagenous colitis including age at onset and clinical presentation, but it has almost equal incidence in men and women and no subepithelial collagen deposition on pathologic section. However, intraepithelial lymphocytes are increased. Diarrhea stops in the majority of patients treated with sulfasalazine or prednisone.

Diversion colitis is an inflammatory process that arises in segments of the large intestine that are excluded from the fecal stream. It usually occurs in patients with ileostomy or colostomy when a mucus fistula or a Hartman's pouch has been created. Diversion colitis is reversible by surgical reanastamosis. Clinically, patients have mucus or bloody discharge from the rectum. Erythema, granularity, friability, and, in more severe cases, ulceration can be seen on endoscopy. Histopathology shows areas of active inflammation with foci of cryptitis and crypt abscesses. Crypt architecture is normal and this differentiates it fromUC. It may be impossible to distinguish fromCD. Short-chain fatty acid enemas will help in diversion colitis, but the definitive therapy is surgical reanastamosis.

EXTRAINTESTINAL MANIFESTATIONS

IBDis associated with a variety of extraintestinal manifestations; up to one-third of patients have at least one. Patients with perianalCD are at higher risk for developing extraintestinal manifestations than other IBD patients.

DERMATOLOGIC

Erythema nodosum (EN) occurs in up to 15% ofCDpatients and 10% ofUCpatients. Attacks usually correlate with bowel activity; skin lesions develop after the onset of bowel symptoms, and patients frequently have concomitant active peripheral arthritis. The lesions of EN are hot, red, tender nodules measuring 1 to 5 cm in diameter and are found on the anterior surface of the lower legs, ankles, calves, thighs, and arms. Therapy is directed toward the underlying bowel disease.

Pyoderma gangrenosum (PG) is seen in 1 to 12% ofUCpatients and less commonly inCDcolitis. Although it usually presents after the diagnosis ofIBD, PG may occur years before the onset of bowel symptoms, run a course independent of the bowel disease, respond poorly to colectomy, and even develop years after proctocolectomy. It is usually associated with severe disease. Lesions are commonly found on the dorsal surface of the feet and legs but may occur on the arms, chest, stoma, and even the face. PG usually begins as a pustule and then spreads concentrically to rapidly undermine healthy skin. Lesions then ulcerate with violaceous edges surrounded by a margin of erythema. Centrally, they contain necrotic tissue with blood and exudates. Lesions may be single or multiple and grow as large as 30 cm. They are sometimes very difficult to treat and often require intravenous antibiotics, intravenous glucocorticoids, dapsone, purinethinol, thalidomide, or intravenous cyclosporine.

Other dermatologic manifestations include pyoderma vegetans that occurs in intertriginous areas, pyostomatitis vegetans that involves the mucous membranes, and metastaticCD, a rare disorder defined by cutaneous granuloma formation. Psoriasis affects 5 to 10% of patients withIBD and is unrelated to bowel activity. Perianal skin tags are found in 75 to 80% of patients with CD, especially those with colonic involvement. Oral mucosal lesions are seen often in CD and rarely inUC and include aphthous stomatitis and "cobblestone" lesions of the buccal mucosa.

RHEUMATOLOGIC

Peripheral arthritis develops in 15 to 20% ofIBDpatients, is more common inCD, and worsens with exacerbations of bowel activity. It is asymmetric, polyarticular, and migratory, and most often affects large joints of the upper and lower extremities. Treatment is directed at reducing bowel inflammation. In severe UC, colectomy frequently cures the arthritis.

Ankylosing spondylitis (AS) occurs in about 10% ofIBDpatients and is more common inCD thanUC. About two-thirds of IBD patients with AS test positive for the HLA-B27 antigen. The activity of AS is not related to bowel activity and does not remit with glucocorticoids or colectomy. It most often affects the spine and pelvis, producing symptoms of diffuse low-back pain, buttock pain, and morning stiffness. The course is continuous and progressive leading to permanent skeletal damage and deformity.

Sacroiliitis is symmetrical, occurs equally inUC andCD, is often asymptomatic, does not correlate with bowel activity, and does not necessarily progress toAS. Other rheumatic manifestations include hypertrophic osteoarthropathy, osteoporosis and osteomalacia secondary to malabsorption of calcium and vitamin D as well as glucocorticoid therapy, pelvic/femoral osteomyelitis, and relapsing polychondritis.

OCULAR

The incidence of ocular complications inIBDpatients is 1 to 10%. The most common are conjunctivitis, anterior uveitis/iritis, and episcleritis. Uveitis is associated with bothUC andCDcolitis, may be found during periods of remission, and may develop in patients following bowel resection. Symptoms include ocular pain, photophobia, blurred vision, and headache. Prompt intervention, sometimes with systemic glucocorticoids, is required to prevent scarring and visual impairment. Episcleritis is a benign disorder that presents with symptoms of mild ocular burning. It occurs in 3 to 4% of IBD patients, more commonly in CD colitis, and is treated with topical glucocorticoids.

HEPATOBILIARY

Hepatic steatosis is detectable in about half of the abnormal liver biopsies from patients withCD andUC; patients usually present with hepatomegaly. Fatty liver usually results from a combination of chronic debilitating illness, malnutrition, and glucocorticoid therapy. Cholelithiasis is more common in CD than UC and occurs in 10 to 35% of patients with ileitis or ileal resection. Gallstone formation is caused by malabsorption of bile acids resulting in depletion of the bile salt pool and the secretion of lithogenic bile.

Primary sclerosing cholangitis (PSC) is characterized by both intrahepatic and extrahepatic bile duct inflammation and fibrosis (Fig. 287-17), frequently leading to biliary cirrhosis and hepatic failure; 1 to 5% of patients withIBD have PSC, but 50 to 75% of patients with PSC have IBD. Although it can be recognized after the diagnosis of IBD, PSC can be detected earlier or even years after proctocolectomy. Most patients have no symptoms at the time of diagnosis; when symptoms are present they consist of fatigue, jaundice, abdominal pain, fever, anorexia, and malaise. Diagnosis is made by endoscopic retrograde cholangiopancreatography (ERCP), which demonstrates multiple bile duct strictures alternating with relatively normal segments. The bile acid ursodeoxycholic acid (ursodiol) may reduce alkaline phosphatase and serum aminotransferase levels, but histologic improvement has been marginal and it has no definitive long-term benefit. Endoscopic stenting may be palliative for cholestasis secondary to bile duct obstruction. Patients with symptomatic disease develop cirrhosis and liver failure over 5 to 10 years and eventually require liver transplantation. Ten percent of PSC patients develop cholangiocarcinoma and cannot be transplanted. Pericholangitis is a subset of PSC found in about 30% of IBD patients; it is confined to small bile ducts and is usually benign.

UROLOGIC

The most frequent genitourinary complications are calculi, ureteral obstruction, and fistulas. The highest frequency of nephrolithiasis (10 to 20%) occurs in patients withCDfollowing small bowel resection or ileostomy. Calcium oxalate stones develop secondary to hyperoxaluria, which results from increased absorption of dietary oxalate. Normally, dietary calcium combines with luminal oxalate to form insoluble calcium oxalate, which is eliminated in the stool. In patients with ileal dysfunction, however, nonabsorbed fatty acids bind calcium and leave oxalate unbound. The unbound oxalate is then delivered to the colon, where it is readily absorbed, especially in the presence of colonic inflammation.

OTHER

The risk of thromboembolic disease increases whenIBDbecomes active, and patients may present with deep vein thrombosis, pulmonary embolism, cerebrovascular accidents, and arterial emboli. Factors responsible for the hypercoagulable state include reactive thrombocytosis, increased levels of fibrinopeptide A, factor V, factor VIII, fibrinogen, accelerated thromboplastin generation, antithrombin III deficiency secondary to increased gut losses or increased catabolism, and free protein S deficiency. A spectrum of vasculitidies involving small, medium, and large vessels has also been observed in IBD patients.

Patients withIBD have an increased prevalence of osteoporosis secondary to vitamin D deficiency, calcium malabsorption, malnutrition, and corticosteroid use. Deficiencies of vitamin B12 and fat-soluble vitamins may occur after ileal resection or with ileal disease.

More common cardiopulmonary manifestations include endocarditis, myocarditis, pleuropericarditis, and interstitial lung disease. A secondary or reactive amyloidosis can occur in patients with long-standingIBD, especially in patients withCD. Amyloid material is deposited systemically and can cause diarrhea, constipation, and renal failure. The

renal disease can be successfully treated with colchicine. Pancreatitis is a rare extra-intestinal manifestation of IBD and results from duodenal fistulas, ampullary CD, gallstones, PSC, drugs such as 6-mercaptopurine or azathioprine, autoimmune pancreatitis, and primary CD of the pancreas.

TREATMENT

5-ASA Agents The mainstay of therapy for mild to moderateUC andCDcolitis is sulfasalazine and the other 5-ASA agents. Sulfasalazine was originally developed to deliver both antibacterial (sulfapyridine) and anti-inflammatory (5-aminosalicylic acid, 5-ASA) therapy into the connective tissues of joints and the colonic mucosa. The molecular structure provides a convenient delivery system to the colon by allowing the intact molecule to pass through the small intestine after only partial absorption, and to be broken down in the colon by bacterial azo reductases that cleave the azo bond linking the sulfa and 5-ASA moieties. Sulfasalazine is effective in inducing and maintaining remission in mild to moderate UC and CD ileocolitis and colitis, but its high rate of side effects limits its use. Although sulfasalazine is more effective at higher doses, at 6 or 8 g/d up to 30% of patients experience allergic reactions or intolerable side effects such as headache, anorexia, nausea, and vomiting that are attributable to the sulfapyridine moiety. Hypersensitivity reactions, independent of sulfapyridine levels, include rash, fever, hepatitis, agranulocytosis, hypersensitivity pneumonitis, pancreatitis, worsening of colitis, and reversible sperm abnormalities. Sulfasalazine can also impair folate absorption and patients should be supplemented with folic acid.

Newer sulfa-free aminosalicylate preparations deliver increased amounts of the pharmacologically active ingredient of sulfasalazine (5-ASA, mesalamine) to the site of active bowel disease while limiting systemic toxicity. 5-ASA may function through inhibition of NF-kB activity. Sulfa-free aminosalicylate formulations include alternative azo-bonded carriers, 5-ASA dimers, pH-dependent tablets, and continuous-release preparations. Each has the same efficacy as sulfasalazine when equimolar concentrations are used. Olsalazine is composed of two 5-ASA radicals linked by an azo bond which is split in the colon by bacterial reduction and two 5-ASA molecules are released. Olsalazine is similar in effectiveness to sulfasalazine in treatingCD andUC, but up to 17% of patients experience non-bloody diarrhea caused by increased secretion of fluid in the small bowel. Balsalazide contains an azo bond binding mesalamine to the carrier molecule 4-amino benzoyl balanine; it is effective in the colon. Claversal is an enteric-coated form of 5-ASA that consists of mesalamine surrounded by an acrylic-based polymer resin and a cellulose coating that releases mesalamine at pH> 6.0, a level that is present from the mid-jejunum continuously to the distal colon.

The most commonly used drugs besides sulfasalazine in the United States are Asacol and Pentasa. Asacol is also an enteric-coated form of mesalamine, but it has a slightly different release pattern, with 5-ASA liberated at pH> 7.0. The disintegration of Asacol is variable with complete break-up of the tablet occurring in many different parts of the gut ranging from the small intestine to the splenic flexure; it has increased gastric residence when taken with a meal. Asacol is used to induce and maintain remission inUC and in CDileitis, ileocolitis, and colitis. Appropriate doses of Asacol and the other 5-ASA compounds are shown in Table 287-6. Some 50 to 75% of patients with mild to moderate UC and CD improve when treated with 2 g/d of 5-ASA; the dose response

continues up to at least 4.8 g/d. Doses of 1.5-4 g/d maintain remission in 50 to 75% of patients with UC and CD.

Pentasa is another mesalamine formulation that uses an ethylcellulose coating to allow water absorption into small beads containing the mesalamine. Water dissolves the 5-ASA, which then diffuses out of the bead into the lumen. Disintegration of the capsule occurs in the stomach. The microspheres then disperse throughout the entire gastrointestinal tract from the small intestine through the distal colon in both fasted or fed conditions. Controlled trials of Pentasa and Asacol in activeCDdemonstrate a 40 to 60% clinical improvement or remission, and meta-analyses demonstrate maintenance of CD remission with 1.5 to 3 g/d of 5-ASA in 68 to 95% of patients. Pentasa at a dose of 2 g/d is more effective than placebo in postoperative prophylaxis of CD.

Topical mesalamine enemas are effective in mild-to-moderate distalUC andCD. Clinical response occurs in up to 80% of UC patients with colitis distal to the splenic flexure. Mesalamine suppositories, which are no longer available in the United States but are available in Canada, at doses of 500 mg twice a day are effective in treating proctitis.

Glucocorticoids The majority of patients with moderate to severeUCbenefit from oral or parenteral glucocorticoids. Prednisone is usually started at doses of 40 to 60 mg/d for active UC that is unresponsive to 5-ASA therapy. Parenteral glucocorticoids may be administered as intravenous hydrocortisone 300 mg/d or methylprednisolone 40 to 60 mg/d. Adrenocorticotropic hormone (ACTH) is occasionally preferred for glucocorticoid-naive patients despite a risk of adrenal hemorrhage. ACTH has equivalent efficacy to intravenous hydrocortisone in both glucocorticoid-naive and -experiencedCDpatients.

Topically applied glucocorticoids are also beneficial for distal colitis and may serve as an adjunct in those who have rectal involvement plus more proximal disease. Hydrocortisone enemas or foam may control active disease, although they have no proven role as maintenance therapy. These glucocorticoids are significantly absorbed from the rectum and can lead to adrenal suppression with prolonged administration. The systemic effects of standard glucocorticoid formulations have led to the development of more potent formulations that are less well absorbed and have increased first-pass metabolism. Budesonide is being used in enema form with favorable preliminary results in distalUC.

Glucocorticoids are also effective for treatment of moderate-to-severeCD and induce a 60 to 70% remission rate compared to a 30% placebo response. Controlled ileal-release budesonide has been nearly equal to prednisone for ileocolonic CD at a dose of 9 mg/d.

Glucocorticoids play no role in maintenance therapy in eitherUC orCD. Once clinical remission has been induced, they should be tapered according to the clinical activity, normally at a rate of no more than 5 mg per week. They can usually be tapered to 20 mg/d within 4 to 5 weeks but often take several months to be discontinued altogether. The side effects are numerous, including fluid retention, abdominal striae, fat redistribution, hyperglycemia, subcapsular cataracts, osteonecrosis, myopathy, emotional disturbances, and withdrawal symptoms. Most of these side effects, aside from osteonecrosis, are related to the dose and duration of therapy.

Antibiotics Despite numerous trials, antibiotics have no role in the treatment of active or quiescentUC. However, pouchitis, which occurs in about a third of UC patients after colectomy and ileal pouch-anal anastamosis, usually responds to treatment with metronidazole or ciprofloxacin.

Metronidazole is effective in active inflammatory, fistulous, and perianalCD and may prevent recurrence after ileal resection. The most effective dose is 15 to 20 mg/kg per day in three divided doses; it is usually continued for several months. Common side effects include nausea, metallic taste, and disulfiram-like reaction. Peripheral neuropathy can occur with prolonged administration (several months) and on rare occasions is permanent despite discontinuation. Ciprofloxacin (500 mg bid) is also beneficial for inflammatory, perianal, and fistulous CD. These two antibiotics should be used as second-line drugs in active CD after 5-ASA agents and as first-line drugs in perianal and fistulous CD.

Azathioprine and 6-Mercaptopurine Azathioprine and 6-mercaptopurine (6-MP) are purine analogues commonly employed in the management of glucocorticoid-dependentIBD. Azathioprine is rapidly absorbed and converted to 6-MP, which is then metabolized to the active end product, thioinosinic acid, an inhibitor of purine ribonucleotide synthesis and cell proliferation. These agents also inhibit the immune response. Efficacy is seen at 3 to 4 weeks. Compliance can be monitored by measuring the level of 6-thioguanine, an end product of 6-MP metabolism. Azathioprine (2.0 to 2.5 mg/kg per day) or 6-MP (1.0-1.5 mg/kg per day) have been employed successfully as glucocorticoid-sparing agents in up to two-thirds of UCand CDpatients previously unable to be weaned from glucocorticoids. The role of these immunomodulators as maintenance therapy in UC and CD and for treating active perianal disease and fistulas in CD appears promising. In addition, 6-MP at a dose of 50 mg/d is more effective than Pentasa or placebo for postoperative prophylaxis of CD.

Although azathioprine and6-MP are usually well tolerated, pancreatitis occurs in 3 to 4% of patients, typically presents within the first few weeks of therapy, and is always completely reversible when the drug is stopped. Other side effects include nausea, fever, rash, and hepatitis. Bone marrow suppression (particularly leukopenia) is dose-related and often delayed, necessitating regular monitoring of the complete blood count. Additionally, 1 in 300 individuals lacks thiopurine methyltransferase, the enzyme responsible for drug metabolism; an additional 11% of the population are heterozygotes with intermediate enzyme activity. Both are at increased risk of toxicity because of increased accumulation of thioguanine metabolites. No increased risk of cancer has been documented inIBDpatients taking these medications long-term.

Methotrexate Methotrexate (MTX) inhibits dihydrofolate reductase, resulting in impaired DNA synthesis. Additional anti-inflammatory properties may be related to decreasedIL-1 production. Intramuscular or subcutaneous MTX (25 mg per week) is effective in inducing remission and reducing glucocorticoid dosage and 15 mg per week is effective in maintaining remission in activeCD. Potential toxicities include leukopenia and hepatic fibrosis, necessitating periodic evaluation of complete blood counts and liver enzymes. The role of liver biopsy in patients on long-term MTX is uncertain. Hypersensitivity pneumonitis is a rare but serious complication of therapy. MTX should only be used

when either6-MP or azathioprine are ineffective or poorly tolerated.

Cyclosporine Cyclosporine (CSA) alters the immune response by acting as a potent inhibitor of T cell-mediated responses. Although CSA acts primarily via inhibition ofIL-2 production from T helper cells, it also decreases recruitment of cytotoxic T cells and blocks other cytokines, including IL-3, IL-4, interferona, andTNF. It has a more rapid onset of action than6-MP and azathioprine.

CSAis most effective given at 4 mg/kg per day IV in severeUC that is refractory to intravenous glucocorticoids, with 82% of patients responding. CSA can be an alternative to colectomy. The long-term success of oral CSA is not as dramatic, but if patients are started on6-MP or azathioprine at the time of hospital discharge, remission can be maintained. Intravenous CSA is effective in 80% of patients with refractory fistulas, but 6-MP or azathioprine must be used to maintain remission. Oral CSA alone is only effective at a higher dose (7.5 mg/kg per day) in active disease but is not effective in maintaining remission without 6-MP/azathioprine. Serum levels should be monitored and kept in the range of 200 to 400 ng/mL.

CSAhas the potential for significant toxicity, and renal function should be frequently monitored. Hypertension, gingival hyperplasia, hypertrichosis, paresthesias, tremors, headaches, and electrolyte abnormalities are common side effects. Creatinine elevation calls for dose reduction or discontinuation. Seizures may also complicate therapy, especially if serum cholesterol levels are less than 120 mg/dL. Opportunistic infections, most notably *Pneumocystis carinii* pneumonia, have occurred with combination immunosuppressive treatment; prophylaxis should then be given.

Nutritional Therapies Dietary antigens may act as stimuli of the mucosal immune response. Patients with activeCDrespond to bowel rest, along with total enteral or total parenteral nutrition (TPN). Bowel rest and TPN are as effective as glucocorticoids for inducing remission of active CD but are not as effective as maintenance therapy. Enteral nutrition in the form of elemental or peptide-based preparations are also as effective as glucocorticiods or TPN, but these diets are not palatable. Enteral diets may provide the small intestine with nutrients vital to cell growth and do not have the complications of TPN. In contrast to CD, activeUC is not effectively treated with either elemental diets or TPN. Standard medical management of UC and CD is reviewed inTable 287-7.

Newer Medical Therapies

Anti-tumor Necrosis Factor Antibody TNF is a key inflammatory cytokine and mediator of intestinal inflammation. The expression of TNF is increased inIBD. Infliximab is a chimeric mouse-human monoclonal antibody against TNF that is extremely effective in CD. It blocks TNF in the serum and at the cell surface and likely lyses TNF-producing macrophages and T cells through complement fixation and antibody-dependent cytotoxicity. Of activeCDpatients refractory to glucocorticoids, 6-MP, or 5-ASA, 65% will respond to intravenous infliximab (5 mg/kg); one-third will enter complete remission. Patients who experience an initial response will respond again to repeated infusions of infliximab every 8 weeks up to 44 weeks. Thus infliximab may be also be efficacious in maintaining remission. However, more trials need to be completed on remission

maintenance after infliximab therapy.

Infliximab is also effective inCDpatients with refractory perianal and enterocutaneous fistulas, with a 68% response rate (50% reduction in fistula drainage) and a 50% complete remission rate. The effects of infliximab for both inflammatory and fistulous disease last 12 weeks on average but longer in some patients.

The incidence of antibodies to infliximab (25% of the molecule is murine) is 13%. One side effect is a lupus-like syndrome, which is rare and reversible after stopping the drug. Anti-double-stranded DNA antibodies occur in 9% but are not associated with clinical lupus.

Among more than 1000 patients treated with infliximab, four developed lymphoma: one patient withCD, two with rheumatoid arthritis, and one with AIDS. Since the risk of lymphoma is already increased in these conditions, it is unclear whether infliximab is the cause. Thus, infliximab is extremely effective in refractory inflammatory and fistulous CD, but should be used only when necessary. Results on the efficacy of infliximab in UC are mixed.

Newer Immunosuppressive Agents Tacrolimus has a mechanism of action similar to cyclosporine. It has shown efficacy in children with refractoryIBD and in adults with extensive involvement of the small bowel.

Mycophenolate mofetil inhibits the de novo pathway of purine synthesis in lymphocytes, disrupting the conversion of inosine monophosphate to guanosine monophosphate (GMP) by reversible inhibition of inosine monophosphate dehydrogenase. The resulting depletion of intracellular GMP suppresses the generation of cytotoxic T cells and formation of antibodies by activated B cells. Patients withCD orUC who received either 500 mg twice a day or 15 mg/kg per day in two divided doses have tolerated the drug well and have experienced benefit with reduction of glucocorticoid requirements.

Thalidomide has been shown to inhibitTNFproduction by monocytes and other cells. Thalidomide is effective in glucocorticoid refractory and fistulousCD, but randomized controlled trials still need to be performed.

The Anti-Inflammatory Cytokines IL-10 is an anti-inflammatory and immunosuppressive cytokine produced by subsets of T and B cells, macrophages, and monocytes. It decreases T_H1 production of IL-2 and interferon g, and limits production of IL-1, IL-6, IL-8,TNF, IL-12, and granulocyte-macrophage colony-stimulating factor. IL-10 has a moderate benefit in activeCD.

IL-11 is a cytokine with thrombopoietic activity and mucosal protective effects that is effective in reducing inflammation in animal models of colitis. It seems to be effective in activeCD, but more trials are needed.

Surgical Therapy

Ulcerative Colitis Nearly half of patients with extensive chronicUCundergo surgery within the first 10 years of their illness. The indications for surgery are listed in Table

287-7. Morbidity is about 20% in elective, 30% for urgent, and 40% for emergency proctocolectomy. The risks are primarily hemorrhage, contamination and sepsis, and neural injury. Although single-stage total proctocolectomy with ileostomy has been the operation of choice, newer operations maintain continence while surgically removing the involved rectal mucosa.

TheIPAA is the most frequent continence-preserving operation performed. BecauseUC is a mucosal disease, the rectal mucosa can be dissected out and removed down to the dentate line of the anus or about 2 cm proximal to it. The ileum is fashioned into a pouch that serves as a neorectum. This ileal pouch is then sutured circumferentially to the anus in an end-to-end fashion. If performed carefully, this operation preserves the anal sphincter and maintains continence. The overall operative morbidity is 10%, with the major complication being bowel obstruction. Pouch failure necessitating conversion to permanent ileostomy occurs in 5 to 10% of patients. Some inflamed rectal mucosa is usually left behind, and thus endoscopic surveillance is necessary. Primary dysplasia of the ileal mucosa of the pouch has occurred rarely.

Patients withIPAAs usually have about six to eight bowel movements a day. On validated quality of life indices, they report better performance in sports and sexual activities than ileostomy patients. The most frequent late complication of IPAA is pouchitis in about one-third of patients withUC. This syndrome consists of increased stool frequency, watery stools, cramping, urgency, nocturnal leakage of stool, arthralgias, malaise, and fever. Although it usually responds to antibiotics, in 3% of patients it is refractory and requires pouch take-down.

Crohn's Disease Most patients withCDrequire at least one operation in their lifetime. The need for surgery is related to duration of disease and the site of involvement. Patients with small bowel disease have an 80% chance of requiring surgery. Those with colitis alone have a 50% chance. The indications for surgery are shown inTable 287-7.

SMALL INTESTINAL DISEASE BecauseCD is chronic and recurrent with no clear surgical cure, as little intestine as possible is resected. Current surgical alternatives for treatment of obstructing CD include resection of the diseased segment and strictureplasty. Surgical resection of the diseased segment is the most frequently performed operation, and in most cases primary anastomosis can be done to restore continuity. If much of the small bowel has already been resected and the strictures are short with intervening areas of normal mucosa, strictureplasties should be done to avoid a functionally insufficient length of bowel. The strictured area of intestine is incised longitudinally and the incision sutured transversely, thus widening the narrowed area. Complications of strictureplasty include prolonged ileus, hemorrhage, fistula, abscess, leak, and restricture.

Colorectal Disease A greater percentage of patients withCDcolitis require surgery for intractability, fulminant disease, and anorectal disease. Several alternatives are available, ranging from the use of a temporary loop ileostomy to resection of segments of diseased colon or even the entire colon and rectum. For patients with segmental involvement, segmental colon resection with primary anastomosis can be performed. In 20 to 25% of patients with extensive colitis, the rectum is spared sufficiently to consider rectal preservation. Most surgeons believe that anIPAA is contraindicated in CD due to

the high incidence of pouch failure. A diverting colostomy may help heal severe perianal disease or rectovaginal fistulas, but disease almost always recurs with reanastomosis. Often, these patients require a total proctocolectomy and ileostomy.

INFLAMMATORY BOWEL DISEASE AND PREGNANCY

When adjusted for patient age, the fertility rate inUC is probably normal. In contrast, fertility is reduced inCD in proportion to disease activity and can be restored when remission is induced. The ovaries and fallopian tubes can be affected by the inflammatory process of CD, especially on the right side because of the proximity of the terminal ileum. In addition, perirectal, perineal, rectovaginal abscesses, and fistulae can result in dyspareunia. Infertility in men can be caused by sulfasalazine but reverses when treatment is stopped.

InUC, fetal outcome approximates that in the normal population. InCD, spontaneous abortions, stillbirths, and developmental defects are increased with increased disease activity, not medications. The courses of CD and UC during pregnancy mostly correlate with disease activity at the time of conception. Most CD patients can deliver vaginally, but cesarean section may be the preferred route of delivery for patients with anorectal and perirectal abscesses and fistulas to reduce the likelihood of fistulas developing or extending into the episiotomy scar.

Sulfasalazine, mesalamine, and olsalazine are safe for use in pregnancy, but folate supplementation must be given with sulfasalazine. No adverse affects have been reported from sulfasalazine in nursing infants. Topical 5-ASA agents are also safe during pregnancy. Glucocorticoids are generally safe for use during pregnancy and are indicated for patients with moderate to severe disease activity. The amount of glucocorticoids received by the nursing infant is minimal. The safest antibiotics to use forCD in pregnancy are ampicillin, cephalosporin, or ciprofloxicin. Flagyl is teratogenic and tumorigenic in high doses, passes into breast milk, and should be avoided.

6-MPand azathioprine pose minimal or no risk during pregnancy, but experience is limited. If the patient cannot be weaned from the drug or has an exacerbation that requires 6-MP/azathioprine during pregnancy, she should continue the drug with informed consent. Their effects during nursing are unknown.

There is little data on cyclosporine in pregnancy. In a small number of patients with severeIBDtreated with intravenous cyclosporine during pregnancy, 80% of pregnancies were successfully completed without development of renal toxicity, congenital malformations, or developmental defects. However, because of the lack of data, cyclosporine should probably be avoided unless the patient would otherwise require surgery. Methotrexate is contraindicated in pregnancy and nursing.

Surgery inUCshould be performed only for emergency indications, including severe hemorrhage, perforation, and megacolon refractory to medical therapy. Total colectomy and ileostomy carries a 60% risk of postoperative spontaneous abortion. Fetal mortality is also high in CD requiring surgery. Patients with ileostomies andIPAAs tolerate pregnancy well.

INFLAMMATORY BOWEL DISEASE IN THE ELDERLY

The most common presenting symptoms in the elderly are diarrhea, weight loss, and abdominal pain. CDin the elderly is mostly colonic with a distal distribution and occurs predominantly in women. Proctitis has been documented in 50% of elderly patients and the diagnosis is often delayed. Diseases that can mimic CD in the elderly are ischemic colitis, diverticular disease, irritable bowel, infectious colitides, and malignancies, including carcinoma, lymphoma, and carcinoid. The incidence of surgery is high in elderly patients, with up to 50% of patients with ileitis, ileocolitis, or extensive colitis requiring urgent or early surgery for first-time disease. In addition, surgery has a much higher morbidity than in younger patients, although the rate of postoperative recurrence is less. Most elderly patients respond as well as younger individuals to medical management.

UCin the elderly is more common in men, presents usually with diarrhea and weight loss, and may have a more distal distribution than in younger patients. Most elderly patients have a favorable response to medical therapy, especially 5-ASA agents, and immunosuppressives used in conjunction with low doses of glucocorticoids. Cyclosporine has been used more frequently in the elderly, but the age-related decreases in renal clearance may affect dosing. Glucocorticoid complications such as osteoporosis and hyperglycemia are also increased in the elderly.6-MP and azathioprine are well tolerated in the elderly. Surgery also has a higher morbidity and mortality in UC, and elderly patients have a longer hospital stay than younger patients. The risk of colon cancer in UC andCDcolitis is no greater than that in the general population since the duration of disease is short and the extent of disease is often distal.

CANCER IN INFLAMMATORY BOWEL DISEASE

ULCERATIVE COLITIS

Patients with long-standingUC are at increased risk for developing colonic epithelial dysplasia and carcinoma (Figs. 287-18,287-19, and Fig. 287-20). Several features distinguish sporadic (SCC) and colitis-associated (CAC) colon cancers. First, SCC usually arise from an adenomatous polyp; CAC typically arise from either flat dysplasia or a dysplasia-associated lesion or mass (DALM). Second, multiple synchronous colon cancers occur in 3 to 5% of SCC but in 12% of CAC. Third, the mean age of individuals with SCC is in the sixties; the mean age of those with CAC is in the thirties. Fourth, SCC exhibits a left-sided predominance, whereas CAC is distributed more uniformly throughout the colon. Fifth, mucinous and anaplastic cancers are more common in CAC than SCC. At the molecular level, p53 mutations occur much earlier and *APC* gene mutations much later in CAC than SCC.

The risk of neoplasia in chronicUCincreases with duration and extent of disease. For patients with pancolitis, the risk of cancer rises 0.5 to 1% per year after 8 to 10 years of disease. This observed increase in cancer rates has led to the endorsement of surveillance colonoscopies with biopsies for patients with chronic UC as the standard of care. Annual or biennial colonoscopy with multiple biopsies has been advocated for patients with more than 8 to 10 years of pancolitis or 12 to 15 years of left-sided colitis and has been widely employed to screen and survey for subsequent dysplasia and

carcinoma.

CROHN'S DISEASE

Risk factors for developing colorectal cancer inCD are a history of colonic (or ileocolonic) involvement and long disease duration. The cancer risks in CD andUC are probably equivalent for similar extent and duration of disease. In patients with extensive colonic involvement, the overall risk is increased 18-fold and the cumulative risk is 8% at 22 years. Thus, the same endoscopic surveillance strategy used for UC is recommended for patients with chronic CD colitis. A pediatric colonoscope can be used to pass narrow strictures in CD patients and impassable strictures can be surveyed with annual barium enemas. A colon resection should be performed if there is evidence of malignancy.

MANAGEMENT OF DYSPLASIA AND CANCER

If high grade dysplasia (HGD) is encountered on colonoscopic surveillance, the usual treatment forUC is colectomy and forCD is either colectomy or segmental resection. If low grade dysplasia (LGD) is found, the management is controversial. Many investigators recommend immediate colectomy, but some repeat the colonoscopy in 1 to 6 months and search for recurrent dysplasia. Polyps in chronic colitis can be removed endoscopically provided that biopsies of the surrounding mucosa are free of dysplasia.

IBDpatients are also at greater risk for other malignancies. Patients withCD may have an increased risk of developing non-Hodgkin's lymphoma and squamous cell carcinoma of the skin. Although CD patients have a twelvefold increased risk of developing small bowel cancer, this type of carcinoma is extremely rare.

QUALITY OF LIFE IN INFLAMMATORY BOWEL DISEASE

The assessment of health-related quality of life plays an important role in the evaluation and treatment ofIBDpatients. Although clinical trials have generally relied upon traditional disease activity indices such as the Crohn's Disease Activity Index (CDAI) to measure therapeutic efficacy, these measures do not reflect quality of life. The Inflammatory Bowel Disease Questionnaire (IBDQ) is a validated, disease-specific instrument that has been used to measure quality of life. It is a 32-item questionnaire that measures global function, systemic and bowel symptoms, functional and social impairment, and emotional function. When compared to the general population, IBD patients have an impaired quality of life in all six categories. The most frequent concerns ofUCpatients are having an ostomy bag, developing cancer, effects of medication, the uncertain nature of the disease, and having surgery. The most frequent concerns of CDpatients are the uncertain nature of the disease, energy level, effects of medication, having surgery, and having an ostomy bag.

(Bibliography omitted in Palm version)

Back to Table of Contents

288. IRRITABLE BOWEL SYNDROME - *Chung Owyang*

Irritable bowel syndrome (IBS) is a gastrointestinal (GI) disorder characterized by altered bowel habits and abdominal pain in the absence of detectable structural abnormalities. No clear diagnostic markers exist for IBS, so all definitions of the disease are based on the clinical presentation. The Rome criteria for the diagnosis of IBS are summarized inTable 288-1. IBS is one of the most common conditions encountered in clinical practice but one of the least well understood. Until recently, many physicians did not consider IBS to be a disease at all; they viewed it as nothing more than a somatic manifestation of psychological stress. With the availability of better techniques to study colonic and GI motility and visceral sensory function, along with the development of newer concepts about the importance of the brain in regulating gut function, significant progress has been made toward a better understanding of the pathogenesis of IBS. Improved methods of treatment may result from these insights.

CLINICAL FEATURES

IBSis a disorder of young people, with most new cases presenting before age 45. However, some reports suggest that the elderly are troubled by IBS symptoms up to 92% as often as middle-aged persons. Indeed, many of the diagnoses of "painful diverticular disease" given the elderly patients may represent IBS. Women are diagnosed with IBS two to three times as often as men and make up 80% of the population with severe IBS. Patients with IBS may fall into two broad clinical groups. Most commonly, patients have abdominal pain associated with altered bowel habits that include constipation, diarrhea, or both. In the second group, patients have painless diarrhea. This latter group accounts for<20% of patients with IBS; their condition may be a separate entity but is generally considered a variant of IBS.

Abdominal Pain Abdominal pain inIBS is highly variable in intensity and location. Pain in IBS is localized to the hypogastrum in 25%, the right side in 20%, to the left side in 20%, and the epigastrum in 10% of patients. Pain is frequently episodic and crampy but may be superimposed on a background of constant ache. Pain may be mild enough to be ignored or it may interfere with daily activities. Despite this, malnutrition due to inadequate caloric intake is exceedingly rare with IBS. Sleep deprivation is also unusual because abdominal pain is almost uniformly present only during waking hours. Pain is often exacerbated by eating or emotional stress and relieved by passage of flatus or stools.

Altered Bowel Habits Alteration in bowel habits is the most consistent clinical feature inIBS. Symptoms usually begin in adult life. The most common pattern is constipation alternating with diarrhea, usually with one of these symptoms predominating. At first, constipation may be episodic, but eventually it becomes continuous and increasingly intractable to treatment with laxatives. Stools are usually hard with narrowed caliber, possibly reflecting excessive dehydration caused by prolonged colonic retention and spasm. Most patients also experience a sense of incomplete evacuation, thus leading to repeated attempts at defecation in a short time span. Patients whose predominant symptom is constipation may have weeks or months of constipation interrupted with brief periods of diarrhea. In other patients, diarrhea may be the predominant symptom. Diarrhea resulting from IBS usually consists of small volumes of loose stools, and most

patients have stool volumes of <200 mL. Nocturnal diarrhea does not occur in IBS. Diarrhea may be aggravated by emotional stress or eating. Stool may be accompanied by passage of large amounts of mucus; hence, the term *mucous colitis* has been used to describe IBS. This is a misnomer, since inflammation is not present. Bleeding is not a feature of IBS unless hemorrhoids are present, and malabsorption or weight loss does not occur.

Gas and Flatulence Patients with IBS frequently complain of abdominal distention and increased belching or flatulence, all of which they attribute to increased gas. Although some patients with these symptoms actually may have a larger amount of gas, quantitative measurements reveal that most patients who complain of increased gas generate no more than a normal amount of intestinal gas. Most IBS patients develop symptoms even with minimal gut distention, suggesting that the basis of their complaints is reduced tolerance of distention rather than an abnormal quantity of intraluminal gas. In addition, patients with IBS tend to reflux gas from the distal to the more proximal intestine, which may explain the belching.

Upper Gastrointestinal Symptoms Between 25 and 50% of patients withIBScomplain of dyspepsia, heartburn, nausea, and vomiting. This suggests that areas of the gut other than the colon may be involved. Prolonged ambulant recordings of small bowel motility in patients with IBS show a high incidence of abnormalities in the small bowel during the waking period; nocturnal motor patterns are no different from those of healthy controls. A characteristic finding is the frequent occurrence of episodes of clustered contractions recurring at 0- to 9-min intervals. These episodes have a mean duration of 46 min and are often associated with transient abdominal pain and discomfort. A similar pattern has been observed in patients with IBS by the application of psychological stressors and by intravenous neostigmine. In addition, temporary abolition of migrating motor complexes is observed in IBS patients under mental stress. Thus, IBS appears to be a paroxysmal motor disorder that may be detected in the small bowel.

PATHOPHYSIOLOGY

The pathogenesis ofIBS is poorly understood, although roles for abnormal gut motor and sensory activity, central neural dysfunction, psychological disturbances, stress, and luminal factors have been proposed.

Colonic myoelectrical and motor activity under unstimulated conditions are generally normal, but abnormalities are more prominent under stimulated conditions inIBS. IBS patients may exhibit increased rectosigmoid motor activity for up to 3 h after eating. Provocative stimuli also induce exaggerated colonic motor responses in IBS patients compared to healthy volunteers. For example, inflation of rectal balloons both in diarrhea- and constipation-predominant IBS patients leads to marked distention-evoked contractile activity, which may be prolonged.

As with studies of motor activity,IBSpatients frequently exhibit exaggerated sensory responses to visceral stimulation. Postprandial pain has been temporally related to entry of food bolus into the cecum in 74% of patients. Exaggerated symptoms can be induced by visceral distention in IBS patients. Rectal balloon inflation produces both nonpainful and painful sensations at lower volumes in IBS patients than in healthy controls without

altering rectal tension, suggestive of visceral afferent dysfunction in IBS. The visceral hyperalgesia of IBS appears to be selective for mechanoreceptor-activated stimuli, as perception of intestinal mucosal electrical stimulation is normal in IBS. Similar studies show gastric and esophageal hypersensitivity in patients with nonulcer dyspepsia and noncardiac chest pain, raising the possibility that these conditions have a similar pathophysiologic basis. In contrast to their enhanced gut sensitivity, IBS patients do not exhibit heightened sensitivity elsewhere in the body. Thus the afferent pathway disturbances in IBS appear to be selective for visceral innervation, with sparing of somatic pathways. The mechanisms responsible for visceral hypersensitivity are unclear. These exaggerated responses may be due to: (1) increased end organ sensitivity with recruitment of "silent" nociceptors; (2) spinal hyperexcitability with activation of nitric oxide and possibly other neurotransmitters; (3) endogenous (cortical and brainstem) modulation of caudad nociceptive transmission; and (4) over time, the possible development of long-term hyperalgesia due to development of neuroplasticity, resulting in permanent or semipermanent changes in neural responses to chronic or recurrent visceral stimulation.

The role of central nervous system (CNS) factors in the pathogenesis ofIBS is strongly suggested by (1) the clinical association of emotional disorders and stress with symptom exacerbation, and (2) the therapeutic response to therapies that act on cerebral cortical sites. Positron emission tomography has shown alterations in regional cerebral blood flow in IBS patients. In healthy individuals, rectal distention increases blood flow in the anterior cingulate cortex, a region with an abundance of opiate receptors, which, when activated, may help to reduce sensory input. In contrast, IBS patients exhibit no increased blood flow in the anterior cingulate gyrus but show activation of the prefrontal cortex, either in response to rectal activation or in anticipation of rectal distention. Activation of the frontal lobes may activate a vigilance network within the brain that increases alertness. The anterior cingulate cortex and the prefrontal cortex appear to have reciprocal inhibitory associations. In patients with IBS, the preferential activation of the prefrontal lobe without activation of the anterior cingulate cortex may represent a form of cerebral dysfunction leading to the increased perception of visceral pain.

Abnormal psychiatric features are recorded in up to 80% ofIBSpatients; however, no single psychiatric diagnosis predominates. An association between prior sexual or physical abuse and development of IBS has been reported. Forms of sexual abuse associated with IBS include verbal aggression, exhibitionism, sexual harassment, sexual touching, and rape. The pathophysiologic relationship between IBS and sexual or physical abuse is unknown. However, physical and sexual abuse may result in hypervigilence to body sensations at theCNSlevel and visceral hypersensitivity at the gut level.

Thus patients withIBSfrequently demonstrate increased motor reactivity of the colon and small bowel to a variety of stimuli and altered visceral sensation associated with lowered sensation thresholds. These may result fromCNS(enteric nervous system) dysregulation.

Approach to the Patient

BecauseIBS is a disorder for which no pathognomonic abnormalities have been identified, its diagnosis relies on recognition of positive clinical features and elimination of other organic diseases. A careful history and physical examination are frequently helpful in establishing the diagnosis. Clinical features suggestive of IBS include the following: recurrence of lower abdominal pain with altered bowel habits over a period of time without progressive deterioration, onset of symptoms during periods of stress or emotional upset, absence of other systemic symptoms such as fever and weight loss, and small-volume stool without any evidence of blood.

On the other hand, the appearance of the disorder for the first time in old age, progressive course from time of onset, persistent diarrhea after a 48-h fast, and presence of nocturnal diarrhea or steatorrheal stools argue against the diagnosis ofIBS.

Because the major symptoms ofIBS -- abdominal pain, abdominal bloating, and alteration in bowel habits -- are common complaints of manyGIorganic disorders, the list of differential diagnoses is long. The quality, location, and timing of pain may be helpful in suggesting specific disorders. Pain due to IBS that occurs in the epigastric or periumbilical area must be differentiated from biliary tract disease, peptic ulcer disorders, intestinal ischemia, and carcinoma of the stomach and pancreas. If pain occurs mainly in the lower abdomen, the possibility of diverticular disease of the colon, inflammatory bowel disease (including ulcerative colitis and Crohn's disease), and carcinoma of the colon must be considered. Postprandial pain accompanied by bloating, nausea, and vomiting suggests gastroparesis or partial intestinal obstruction. Intestinal infestation with *Giardia lamblia* or other parasites may cause similar symptoms. When diarrhea is the major complaint, the possibility of lactase deficiency, laxative abuse, malabsorption, hyperthyroidism, inflammatory bowel disease, and infectious diarrhea must be ruled out. On the other hand, constipation may be a side effect of many different drugs, such as anticholinergic, antihypertensive, and antidepressant medications. Endocrinopathies such as hypothyroidism and hypoparathyroidism must also be considered in the differential diagnosis of constipation, particularly if other systemic signs or symptoms of these endocrinopathies are present. In addition, acute intermittent porphyria and lead poisoning may present in a fashion similar to IBS, with painful constipation as the major complaint. These possibilities are suspected on the basis of their clinical presentations and are confirmed by appropriate serum and urine tests.

BecauseIBS is in part a diagnosis of exclusion, certain diagnostic tests should be performed routinely; others may be required depending on the specific presenting symptoms. Factors to be considered when determining the aggressiveness of the diagnostic evaluation include the duration of symptoms, the change in symptoms over time, the age and sex of the patient, the referral status of the patient, prior diagnostic studies, a family history of colorectal malignancy, and the degree of psychosocial dysfunction. Thus a younger individual with mild symptoms requires a minimal diagnostic evaluation, while an older person or an individual with rapidly progressive symptoms should undergo a more thorough exclusion of organic disease. In general most patients should have a complete blood count and sigmoidoscopic examination; in addition, stool specimens should be examined for ova and parasites. In those >40 years, an air-contrast barium enema or colonoscopy should also be done. In patients whose main symptoms are diarrhea and increased gas, the possibility of lactase

deficiency should be ruled out with a hydrogen breath test or a lactose-free diet should be prescribed for 3 weeks. In patients with concurrent symptoms of dyspepsia, upperGIradiographs or esophagogastroduodenoscopy may be advisable. In patients with postprandial right upper quadrant pain, ultrasound of the gallbladder should be obtained. Laboratory features that argue against IBS include evidence of anemia, elevated sedimentation rate, presence of leukocytes or blood in stool, and stool volume >200 to 300 mL/d. These findings suggest other diagnostic considerations.

TREATMENT

Patient Counseling and Dietary Alterations Reassurance and careful explanation of the functional nature of the disorder and of how to avoid obvious food precipitants are important first steps in patient counseling and dietary change. Occasionally, a meticulous dietary history may reveal substances (such as coffee, disaccharides, legumes, and cabbage) that aggravate symptoms. As a therapeutic trial, patients should be encouraged to eliminate any foodstuffs that appear to produce symptoms.

Stool Bulking Agents High-fiber diets and bulking agents, such as bran or hydrophilic colloid, are frequently used in treatingIBS. Dietary fiber has multiple effects on colonic physiology. The water-holding action of fibers may contribute to increased stool bulk. Fiber also speeds up colonic transit in most people. In diarrhea-prone patients, whole-colonic transit is faster than average; however, dietary fiber can delay transit. Furthermore, because of their hydrophilic properties, stool-bulking agents bind water and thus prevent both excessive hydration or dehydration of stool. A high-fiber diet relieves diarrhea in some IBS patients. Dietary fiber has also been shown to lower pressures in the sigmoid colon in IBS patients. The effects of fiber on pressure in the rest of the colon are unknown; however, the whole colon is affected by IBS, and the pain of colon spasm often originates from the ascending and transverse segments. Fiber supplementation with psyllium reduces the perception of rectal distention, indicating that fiber may have an effect on visceral afferent function.

The beneficial effects of dietary fiber on colonic physiology suggest that dietary fiber should be an effective treatment forIBSpatients, but controlled trials of dietary fiber have produced variable results. IBS is not purely a colonic disorder; many patients may have symptoms originating from the upper gut. Despite the equivocal data regarding efficacy, most gastroenterologists consider stool-bulking agents worth trying in patients with IBS. Patients should be advised to take increasing quantities of bran supplements, such as whole-meal bread, high-bran cereal, or raw bran, until they are passing one or two soft stools daily. Alternatively, psyllium preparations may be used. About 20% of patients, however, complain that a high-fiber diet aggravates such symptoms as bloating and distention. These undesirable effects usually disappear spontaneously after several weeks.

Antispasmodics Anticholinergic drugs may provide temporary relief for symptoms such as painful cramps related to intestinal spasm. Although controlled clinical trials have produced mixed results, evidence generally supports use of anticholinergic drugs for pain. Meta-analysis of 26 double-blind clinical trials of antispasmodic agents inIBSshowed better global improvement (62%) and abdominal pain reduction (64%) compared to placebo (35% and 45%, respectively). The drugs are most effective when

prescribed in anticipation of predictable pain. Physiologic studies demonstrate that anticholinergic drugs inhibit the gastrocolic reflex; hence, postprandial pain is best managed by giving antispasmodics 30 min before meals so that effective blood levels are achieved shortly before the anticipated onset of pain. Most anticholinergics contain natural belladonna alkaloids, which may cause xerostomia, urinary hesitancy and retention, blurred vision, and drowsiness. Some physicians prefer to use synthetic anticholinergics, such as dicyclomine, that have less effect on mucous membrane secretions and therefore produce fewer undesirable side effects.

Antidiarrheal Agents When diarrhea is severe, especially in the painless diarrhea variant ofIBS, small doses of diphenoxylate (Lomotil), 2.5 to 5 mg every 4 to 6 h, can be prescribed. These agents are less addictive than paregoric, codeine, or tincture of opium. In general, the intestines do not become tolerant of the antidiarrheal effect of opiates, and increasing doses are not required to maintain antidiarrheal potency. These agents are most useful if taken before anticipated stressful events that are known to cause diarrhea. Treatment with antidiarrheals, however, should be considered only as temporary management; the final goal of treatment is gradual withdrawal of medication with substitution of a high-fiber diet.

Drug Antidepressants In addition to their mood-elevating effects, antidepressent medications have several physiologic effects that may be beneficial inIBS. In diarrhea-predominant IBS patients, the tricyclic antidepressant imipramine slows jejunal migrating motor complex transit propagation and delays orocecal and whole-gut transit, indicative of a motor inhibitory effect. Tricyclic agents may alter visceral afferent neural function.

Tricyclic antidepressants may be effective in someIBSpatients. In a 2-month study of desipramine, abdominal pain improved in 86% of patients compared to 59% given a placebo. Another study of desipramine in 28 IBS patients showed improvement in stool frequency, diarrhea, pain, and depression. Improvements were mainly observed in diarrhea-predominant patients, with no improvement noted in constipated patients. The efficacy of other antidepressant agents is less well evaluated. An uncontrolled review of antidepressant therapy in 138 patients with IBS, including both tricyclic agents and the newer selective serotonin reuptake inhibitors (e.g., fluxetine, paroxetine, and sertraline), reported symptomatic improvement in 89% of individuals, especially those in the pain-predominant subtype. However, no placebo-controlled trials of the selective serotonin inhibitors have been reported in IBS to date.

Antiflatulence Therapy The management of excessive gas is seldom satisfactory, except in cases of obvious aerophagia or disaccharidase deficiency. Patients should be advised to eat slowly; not chew gum or drink carbonated beverages; and avoid artificial sweeteners, legumes, and foods of the cabbage family. Simethicone, antacids, and activated charcoal have all been tried, usually with disappointing results.

FUTURE DIRECTIONS IN MEDICAL TREATMENT OF IBS

Medications that blunt the visceral hyperalgesia ofIBS are in development. Such "antiafferent" agents might act via one or more mechanisms, including (1) modification of release of pain-inducing mediators in the gut wall, (2) blockade or activation of

peripheral afferent nerve receptors, (3) inhibition of afferent nerve transmission, or (4) modification of afferent activity in theCNS. These include the kappa opioid compounds and serotonin receptor ($5HT_3$) antagonists such as alosetron and octreotide.

Such compounds have been shown to reduce perception of painful mechanical visceral stimulation in patients withIBS. Furthermore, placebo-controlled trials with IBS patients have shown that alosetron or fedotozine, a kappa opioid analogue, reduces both pain and the severity of disease. Additional clinical studies of this group of compounds may lead to new therapeutic approaches for the treatment of IBS.

(Bibliography omitted in Palm version)

Back to Table of Contents

289. DIVERTICULAR, VASCULAR, AND OTHER DISORDERS OF THE INTESTINE AND PERITONEUM - *Kurt J. Isselbacher, Alan Epstein*

DIVERTICULAR DISEASE

Diverticula may be either congenital or acquired and may affect either the small or large intestine. Congenital diverticula are herniations of the entire thickness of intestinal wall, while the more common acquired diverticula consist of herniations of the mucosa through the muscularis, generally at the site of a nutrient artery.

SMALL-INTESTINAL DIVERTICULA

Diverticula may occur in any portion of the small intestine; however, with the exception of Meckel's diverticulum, the most common locations are in the duodenum and jejunum. Most often diverticula are asymptomatic and discovered incidentally on upper gastrointestinal x-rays. On occasion, however, they may cause symptoms either because of their anatomic proximity to other structures or rarely from inflammation or bleeding.

Duodenal diverticula arise singly from the medial surface of the second portion of the duodenum. In most patients they cause no symptoms. Rarely, they may present as acute diverticulitis with abdominal pain, fever, gastrointestinal bleeding, or, most rarely, perforation. Periampullary diverticula are occasionally associated with cholangitis or pancreatitis. Jejunal diverticula, while less common, may also be the site of acute inflammation, bleeding, or perforation with resulting abscess or peritonitis.

Multiple jejunal diverticula may be associated with malabsorption related to bacterial overgrowth within the diverticula, similar to other situations where intestinal stasis (e.g., blind loops) permits bacterial proliferation.**The consequences of bacterial proliferation with resultant mucosal damage, deconjugation of bile salts, and vitamin B_{12} malabsorption are discussed in Chap. 286.*

Meckel's diverticulum, a persistent omphalomesenteric duct, is the most frequent congenital anomaly of the digestive tract, occurring in ~2% of autopsied adults. The diverticulum is wide-mouthed, about 5 cm long, and arises from the antimesenteric border of the ileum, usually within 100 cm of the ileocecal valve. The sac may be lined with normal ileal mucosa (approximately half) or contain gastric, duodenal, pancreatic, or colonic mucosa. While rarely symptomatic after age 5, Meckel's diverticulum may produce hemorrhage, inflammation, and obstruction in children and teenagers.

Hemorrhage occurs almost exclusively before age 10 and invariably results from peptic ulceration of ileal mucosa adjacent to a Meckel's diverticulum lined with gastric mucosa. The diagnosis may be established by isotope scanning of the abdomen after injection of technetium-99, which is taken up by the ectopic gastric mucosa in the diverticulum. False-negative and false-positive Meckel's scans are not uncommon; thus, other clinical and laboratory features must be assessed carefully before recommending surgery. In older children and young adults, inflammation of the diverticulum may mimic acute appendicitis. Mechanical obstruction may also occur if the diverticulum intussuscepts into the lumen of the bowel or twists on a fibrous remnant of the omphalomesenteric

duct that extends from the diverticulum to the abdominal wall. The treatment of any of these complications of Meckel's diverticulum is surgical excision.

COLONIC DIVERTICULA

Diverticula of the colon are herniations or saclike protrusions of the mucosa through the muscularis, at the point where a nutrient artery penetrates the muscularis. Diverticula occur most commonly in the sigmoid colon and decrease in frequency in the proximal colon. They increase with age; the incidence is 20 to 50% in western populations over age 50. The exact mechanism for their formation is unknown but may be related to an increase in intraluminal pressure. Thickening of the muscle coat of the colon in most patients with diverticula suggests that herniations of mucosa are caused by increased pressure produced by colonic muscle contractions. The rarity of colonic diverticula in underdeveloped nations has led to the speculation that diverticula result from the highly refined western diet, which is deficient in dietary fiber or roughage. It is proposed that such diets result in decreased fecal bulk, narrowing of the colon, and an increase in intraluminal pressure in order to move the smaller fecal mass. However, the role of dietary fiber in the etiology and treatment of diverticular disease remains to be determined.

Colonic diverticula are usually asymptomatic and are an incidental finding on barium enema or colonoscopy. The major complications of inflammation, both acute and chronic, and hemorrhage occur in only a small percentage of individuals with diverticulosis. Since diverticulosis is quite common in older patients, one must avoid the temptation of attributing pain or bleeding to the diverticula unless other conditions, especially colon cancer, have been excluded.

DIVERTICULITIS

Inflammation can occur in or around the diverticular sac. The cause of diverticulitis is probably mechanical, related to retention in the diverticula of undigested food residue and bacteria, which may form a hard mass called a *fecalith*. This compromises the blood supply to the thin-walled sac (made up solely of mucosa and serosa) and renders it susceptible to invasion by colonic bacteria. The inflammatory process may vary from a small intramural or pericolic abscess to generalized peritonitis. Some attacks are accompanied by minimal symptoms and seem to heal spontaneously. Studies of resected specimens indicate that most perforations of the diverticular sac are small and result in inflammation of the sac itself and the adjacent serosal surface. Diverticulitis occurs more often in men than in women and three times as often in the left as in the right colon. This suggests that diverticulitis may be related to the higher intraluminal pressures and the more solid fecal material in the sigmoid and descending colon.

Acute colonic diverticulitis is a disease of variable severity characterized by fever, left lower quadrant abdominal pain, and signs of peritoneal irritation -- muscle spasm, guarding, rebound tenderness. Rectal examination may reveal a tender mass if the area of inflammation is close to the rectum. Although constipation may not have been noted before onset of the illness, the inflammation around the colon often results in some degree of acute constipation or obstipation. Rectal bleeding, usually microscopic, is noted in 25% of cases; it is rarely massive. Polymorphonuclear leukocytosis is common.

Complications include free perforation, which results in acute peritonitis, sepsis, and shock, particularly in the elderly. The perforation may be walled off by adherent omentum or neighboring structures such as the bladder or small bowel. Abscess formation or fistulas then occur as the inflammatory mass burrows into other organs. Severe pericolitis may cause a fibrous stricture around the bowel, which can be associated with colonic obstruction and may mimic a neoplasm.

Diagnosis During the acute phase of diverticulitis, barium enema and sigmoidoscopy may be hazardous, since contrast material or air under pressure may lead to rupture of an inflamed diverticulum and convert a walled-off inflammatory lesion to a free perforation. These examinations are usually safe after adequate treatment and healing of the diverticulitis. The radiologic findings on barium enema suggestive of diverticulitis are leakage of barium from a diverticular sac, stricture formation, and the presence of a pericolic inflammatory mass. In many patients the distortion caused by inflammation prevents a clear distinction between cancer and diverticulitis. In these cases, colonoscopy or surgical excision may be required for accurate diagnosis. Abdominal computed tomography scan may demonstrate the presence of a pericolic abscess.

TREATMENT

Most patients with acute diverticulitis require bowel rest, intravenous fluids, and broad-spectrum antibiotics. Repeated attacks of diverticulitis in the same area generally require surgical resection. Severe attacks with acute peritoneal signs, suspected abscess, or perforation require intravenous antibiotics directed against gram-negative anaerobic bacteria, followed by surgical drainage or resection. The usual procedure is a diverting colostomy with resection of the involved colon; reanastomosis is then performed at a second operation.

PAINFUL DIVERTICULAR DISEASE WITHOUT DIVERTICULITIS

Some patients with diverticulosis develop recurrent left lower quadrant colicky pain without clinical or pathologic evidence of acute diverticulitis. They often have bouts of alternating constipation and diarrhea; the pain may be relieved by defecation or passage of flatus. These features suggest the coexistence of the irritable bowel syndrome. Examination during a bout of pain reveals tenderness of the sigmoid colon, but signs of peritoneal inflammation such as rebound tenderness, muscle guarding, fever, and leukocytosis are absent. Barium enema shows typical diverticula without evidence of inflammation and stricture, plus a "sawtooth" irregularity of the lumen, reflecting muscle hypertrophy and spasm. In some patients the pain is severe enough to warrant observation in a hospital and restriction of food, since feeding aggravates the pain by causing colonic contraction. Anticholinergics, which reduce sigmoid contractions, and mild sedation are usually all that is required. After recovery, the patient should be started on a high-residue diet or given a bulk laxative such as hemicellulose, unprocessed bran, or psyllium extract. Surgical excision is usually not indicated unless acute diverticulitis or its complications occur.

HEMORRHAGE FROM DIVERTICULA

Massive hemorrhage from colonic diverticula is one of the most common causes of

hematochezia in patients over age 60. This complication of diverticulosis is caused by erosion of a vessel by a fecalith within the diverticular sac. The bleeding is painless and not accompanied by signs or symptoms of diverticulitis. Most cases of mild or moderate hemorrhage stop spontaneously with bed rest and blood transfusion. Localization of bleeding can be obtained by bleeding scan or angiography. In patients with severe hemorrhage, mesenteric angiography can be both diagnostic in localizing the bleeding site and therapeutic, since vasoconstrictive drugs or artificial blood clot infused intraarterially can sometimes effectively control hemorrhage. Colonoscopy is also useful in evaluating acute hematochezia, and the endoscopist may be able to cauterize angiodysplasias (Chap. 44). The location of bleeding diverticula is more commonly in the right colon, particularly the ascending colon, in contrast to the sigmoid colon, where diverticula are more numerous.

MOTILITY DISORDERS

Normal intestinal motility involves the delicate interplay of the gut motor system, neural influences of the autonomic and central nervous system, as well as hormonal factors, specifically gut neuropeptides. In addition, many drugs used in the treatment of disease (e.g., opioids, antibiotics) affect and influence intestinal motility directly or indirectly. Table 289-1 lists some of the disorders of the enteric motor and neural system. Only the more clinically relevant ones are discussed.

MEGACOLON

Megacolon, or giant colon, is characterized by massive distention of the colon, usually accompanied by severe constipation or obstipation. This condition can be either congenital or acquired and is seen in all age groups. Acute toxic megacolon is a severe complication of ulcerative colitis (Chap. 287).

Aganglionic Megacolon (Hirschsprung's Disease) This is a congenital disorder due to absence of enteric neurons (ganglions) in the distal colon and rectum. This aganglionic segment loses its neural inhibition and remains contracted. Hirschsprung's disease is a heterogeneous genetic disorder -- some patients have an autosomal dominant form of the disease with mutations in the *RET* gene; many have an autosomal recessive form with a mutation in the endothelin-B receptor gene. Hirschsprung's disease is a multigenic trait; genes on 9q31 affect the phenotype of the *RET* gene mutations. These defects result in the gestational failure of neural crest cells to migrate to the distal colon. The disease manifests in early infancy, occurring more frequently in males, and is often familial. These infants have massive abdominal distention, absent bowel movements, and impaired nutrition due to chronic obstruction of the colon. In some individuals with less severe symptoms, the disease may not be diagnosed until adolescence or early adulthood. The aganglionic and contracted segment of bowel is unable to relax to permit passage of stool, causing the normal proximal colon to become greatly dilated. On rectal examination the ampulla is empty of feces and the anal sphincter is normal. Barium enema reveals a narrowed segment in the rectosigmoid, with massive dilation above. Diagnosis is made by full-thickness surgical biopsy under anesthesia and demonstration of absent ganglion cells in the diseased segment. In most patients the aganglionic segment is in the rectosigmoid colon. The treatment of choice is a pull-through procedure in which normally innervated colon is anastomosed to

the distal rectum just above the internal sphincter, thus bypassing the contracted aganglionic segment and restoring normal defecation.

Acquired Megacolon In Central and South America, infection with *Trypanosoma cruzi* (Chagas' disease) can result in destruction of the ganglion cells of the colon, producing a clinical picture similar to congenital megacolon, except that the onset is in adult life rather than in childhood. A number of other diseases are associated with megacolon in adults. Patients with schizophrenia or depression, particularly institutionalized patients, may have obstipation and massive colonic dilatation. Severe neurologic disorders, including cerebral atrophy, spinal cord injury, and parkinsonism, also may cause megacolon. Myxedema, infiltrative diseases such as amyloidosis, and primary systemic sclerosis also can reduce colonic motility and produce marked colonic distention. Narcotic drugs, particularly morphine and codeine, can cause severe constipation, especially when administered to bedridden patients. Digital rectal examination of adults with acquired megacolon reveals a rectum distended with feces, as opposed to the empty rectum in aganglionic megacolon. Treatment is aimed at the underlying disease, as well as the careful use of enemas and cathartics.

Intestinal Pseudoobstruction Intestinal pseudoobstruction is an acute or chronic motility disorder characterized by distention or dilation of the small and large intestine. Abdominal pain, nausea, and vomiting may lead to diagnostic confusion with mechanical obstruction; but as the name of this condition implies, the underlying cause is not obstruction but rather a severe dysmotility resulting in distention. Pseudoobstruction may be primary or secondary and acute or chronic. In primary or idiopathic pseudoobstruction no other contributing condition can be identified, and the motility disorder is attributed to abnormalities of sympathetic innervation or of the muscle layers of the intestine. Secondary pseudoobstruction may result from primary systemic sclerosis, diabetes, amyloidosis, neurologic diseases, drugs, or sepsis.

Chronic or Intermittent Secondary Pseudoobstruction Numerous medical conditions can cause chronic dilation of the large and small bowel. Some of these may involve the intestinal smooth muscle, such as primary systemic sclerosis, amyloidosis, or muscular dystrophy. Endocrine disorders, including myxedema and diabetes mellitus, may result in chronic distention, which in the diabetic patient results from autonomic visceral neuropathy. Chronic neurologic diseases, including Parkinson's disease and stroke, may be complicated by chronic pseudoobstruction; in these patients drugs and relative immobility are contributing features. Finally, institutionalized psychotic patients may suffer from prolonged megacolon.

The symptoms of chronic secondary pseudoobstruction are chronic or intermittent constipation, crampy abdominal pain, anorexia, and bloating. Gastric distention and disordered swallowing may be present. Abdominal x-rays reveal gaseous distention of the large and small bowel and occasionally of the stomach. Air-fluid levels are unusual and should raise the possibility of mechanical obstruction. Upper gastrointestinal series and barium enema do not reveal specific abnormalities of the intestine such as tumor, stricture, or volvulus. The presence of an autoimmune disorder or endocrinopathy may require confirmation by serologic or blood tests; biopsy may be needed as in amyloidosis or muscular dystrophy.

The treatment of chronic intestinal pseudoobstruction is made difficult by the complexity and chronicity of the underlying systemic disease. Patients with primary systemic sclerosis may respond to broad-spectrum antibiotics if intestinal bacterial overgrowth is suspected. Metoclopramide may benefit gastric dysmotility in the diabetic patient. Discontinuation of psychotropic or anti-Parkinson drugs may occasionally result in improvement. Cathartics and enemas may be required to relieve fecal impaction, and the regular use of stool softeners and a high-fiber diet may help prevent recurrences.

Idiopathic Intestinal Pseudoobstruction This term describes the condition of patients with signs and symptoms of pseudoobstruction in whom no systemic disease can be identified. The typical patient has recurrent attacks of abdominal pain and distention with nausea and vomiting. The small intestine is primarily involved, and chronic constipation is much less frequent than in secondary pseudoobstruction. Steatorrhea secondary to bacterial overgrowth of the small intestine is common and may lead to chronic diarrhea and malnutrition. Many patients exhibit abnormalities of motility in the esophagus and urinary bladder, in addition to the small and large intestine. Neuromuscular defects have been described in patients with this syndrome, including abnormalities of the mesenteric plexus and myopathy of the intestinal and urinary bladder smooth muscle (so-called hollow visceral myopathy). Elevated prostaglandin E levels have been reported in some patients.

Management of idiopathic pseudoobstruction is unsatisfactory. Surgery to relieve "obstruction" is to be avoided, since the condition is often worsened by abdominal surgery. Medical therapy with metoclopramide and cholinergic agents has been unsuccessful. Nutritional support in the form of low-residue elemental diets or parenteral hyperalimentation may be helpful. Unfortunately, the lack of effective therapy and the progressive nature of the illness make the prognosis of idiopathic pseudoobstruction rather poor. Death from malnutrition and steatorrhea are common. The long-term impact of total parenteral nutrition on this disease is not yet clear.

Acute Intestinal Pseudoobstruction This entity, sometimes referred to as *Ogilvie's syndrome*, is characterized by acute intestinal dilation involving primarily the colon but occasionally also the small intestine. As in other forms of pseudoobstruction, the clinical features are difficult to distinguish from mechanical obstruction. The patient may complain of colicky lower abdominal pain and acute constipation. Examination reveals a distended, tympanitic abdomen, with reduced or absent bowel sounds. Localized tenderness over the distended colon is common, but diffuse abdominal tenderness, rigidity, or rebound tenderness are unusual. Abdominal films reveal massive dilation of the colon and small intestine, occasionally with the presence of air-fluid levels. The cecum, being the most capacious part of the colon, is often massively dilated and tender. The onset of these symptoms usually occurs in patients who have recently undergone severe surgical or medical stress, such as major surgery, myocardial infarction, sepsis, or respiratory failure. Patients with acute pseudoobstruction are frequently on respirators, have received narcotics or sedatives, and have metabolic and electrolyte disturbances. Ogilvie's syndrome may also be due to paraneoplastic obstruction.

Management of acute pseudoobstruction requires careful correction of fluid and electrolyte abnormalities, intubation of the stomach or small intestine for

decompression, and avoidance of drugs that depress intestinal motility. Barium enema may be hazardous because of the risk of perforating the already dilated bowel. Decompressive colonoscopy is beneficial in some patients, and cecostomy may be required in some patients with massive cecal dilation. The outcome depends in large part on the prognosis of the associated medical or surgical conditions. Patients who recover from the underlying medical or surgical conditions usually have a return of normal colonic function.

IRRITABLE BOWEL SYNDROME

SeeChap. 288.

CHRONIC CONSTIPATION

Chronic constipation is widespread in western society, with ~10% of the population taking laxatives on a regular basis. Most cases of chronic constipation arise from habitual neglect of afferent impulses, failure to initiate defecation, and accumulation of large, dry fecal masses in the rectum. This voluntary suppression of the call to stool may arise during the period of toilet training in childhood or later in life because of a sense of social impropriety, unaccustomed surroundings, uncomfortable toilet facilities, or illnesses that require confinement to bed. Chronic constipation is much more common in women, with onset typically in late adolescence or early adulthood. As constant distention of the rectum with feces becomes chronic, the patient grows less aware of rectal fullness. Bowel movements become progressively more difficult, and painful hemorrhoids or anal fissures reinforce suppression of the urge to defecate. To avoid these problems the patient begins the chronic use of laxatives or enemas, without which defecation becomes impossible.**The mechanism of defecation is discussed in Chap. 42.*

TREATMENT

The physician should make every attempt to educate the patient about the chain of events that has led to chronic constipation. Attempts should be made to alter patterns of many years' duration, and the patient must recognize the importance of responding to, rather than suppressing, the urge to defecate. Defecation should be attempted at a given time each day. In most individuals the call to stool occurs in the morning after breakfast. Physical exercise such as a brisk walk just before attempts at defecation may be helpful. Patients are instructed to increase dietary bulk with foods rich in fiber, such as green vegetables and unprocessed cereal grains, or by the regular use of bulk laxatives, such as hemicellulose, psyllium extract, and powdered unprocessed bran. The success of such a regimen depends to some extent on the duration of symptoms. Elderly patients with long-standing constipation and reliance on enemas or laxatives are more resistant to these measures than younger patients whose bowel patterns are less established. Moreover, poor muscle tone, reduced physical activity, and increased incidence of other medical conditions make the problem more difficult in the older age group. Bedridden elderly patients often develop severe constipation and even fecal impaction unless preventive measures are taken. This applies not only to patients with previous constipation but also to those with regular bowel movements before their confining illness. Regular administration of stool softeners, bulk laxatives, or mild

cathartics is necessary until full ambulation and a normal diet are resumed. The onset of fecal impaction in bedridden patients is heralded by a feeling of rectal distention, urgency of defecation, or tenesmus. Occasionally, the fecal impaction will result in low-grade chronic obstruction with dilation and increased fluid content proximal to the impaction; "paradoxical diarrhea" may thus occur as fluid moves past the obstructing fecal mass. This situation will be aggravated if antidiarrheal drugs are given because the underlying constipation will be worsened. The appropriate maneuver is to disimpact the rectum manually or to administer gentle enemas if the impaction is beyond the reach of the finger.

DISORDERS OF THE MESENTERIC CIRCULATION

Ischemia of the intestine is the end result of interruption or reduction of its blood supply. However, the clinical manifestations of intestinal ischemia range from mild chronic symptoms to a catastrophic acute episode, depending on the vascular supply involved, the extent of the occlusion or ischemia, and the rapidity of the process. The clinician should be aware of the spectrum of clinical manifestations (Table 289-2). The gut derives its arterial blood supply from the celiac axis and the superior and inferior mesenteric arteries. The small intestine is supplied by the celiac and superior mesenteric arteries, the colon by branches of the superior and inferior mesenteric arteries. A rich network of anastomotic vessels and the possible development of collateral circulation determine the clinical picture of acute or chronic intestinal arterial insufficiency.

MESENTERIC ISCHEMIA AND INFARCTION

Acute intestinal ischemia may be classified as occlusive or nonocclusive. *Occlusion* accounts for about 75% of acute intestinal ischemia and may result from an arterial thrombus (one-third of arterial occlusions) or embolus (two-thirds of arterial occlusions) of the celiac or superior mesenteric arteries, or from venous occlusion (<5% of occlusions) in the same distribution. Arterial embolus occurs most commonly in patients with chronic or recurrent atrial fibrillation, artificial heart valves, or valvular heart disease; arterial thrombosis is usually associated with extensive atherosclerosis or low cardiac output. Venous occlusion is rare; it is occasionally seen in women taking oral contraceptives. Approximately one-fourth of patients with mesenteric ischemia have no definite occlusion of a major vessel, a condition referred to as *nonocclusive ischemia*. The exact cause of nonocclusive disease is obscure; systemic arterial hypotension, cardiac arrhythmias, prolonged heart failure, digitalis therapy, dehydration, and endotoxemia can be contributing factors.

The major clinical feature of acute mesenteric ischemia is severe abdominal pain, often colicky and periumbilical at the onset, later becoming diffuse and constant. Vomiting, anorexia, diarrhea, and constipation are also frequent but of little diagnostic help. Examination of the abdomen may reveal tenderness and distention. Bowel sounds are often normal even in the face of severe infarction. Some patients have a surprisingly normal abdominal examination in spite of severe pain. Mild gastrointestinal bleeding is often detected by examination of stool for occult blood; gross hemorrhage is unusual except in ischemic colitis. Leukocytosis is often present. Late in the course of the disease (24 to 72 h), gangrene of the bowel occurs with diffuse peritonitis, sepsis, and

shock. Abdominal plain films in patients with mesenteric ischemia may reveal air-fluid levels and distention. Barium study of the small intestine reveals nonspecific dilation, poor motility, and evidence of thick mucosal folds ("thumbprinting") (Fig. 289-1).

Acute mesenteric ischemia is a grave condition with a high morbidity and mortality. Patients suspected of having acute arterial embolus should undergo immediate celiac and mesenteric angiography to localize the embolus, followed by embolectomy. Restoration of normal circulation may allow complete recovery if performed before irreversible necrosis or gangrene has occurred. Unfortunately, infarction and transmural necrosis are frequently found at surgery, necessitating resection. Arterial or venous thrombosis is not generally amenable to surgical removal of the thrombus, and resection of the affected bowel is required. Similarly, patients with nonocclusive ischemia are not candidates for corrective vascular surgery (as major vessels are patent). These individuals often have extensive necrosis of the small or large intestine because of the widespread nature of the ischemic event. The decision to operate when mesenteric ischemia is suspected is often difficult, because the typical patient is a poor surgical risk owing to advanced age, dehydration, sepsis, and other serious medical conditions.

Chronic arterial insufficiency may precede acute vascular insufficiency, producing so-called abdominal angina. As in angina pectoris, the pain of chronic mesenteric insufficiency occurs under conditions of increased demand for splanchnic blood flow. The patient complains of intermittent dull or cramping midabdominal pain 15 to 30 min after a meal, lasting for several hours postprandially. Significant weight loss due to decreased food intake may be present. Chronic intestinal ischemia also may produce mucosal damage and malabsorption, which in turn aggravates the weight loss. Since abdominal angina may progress to bowel infarction, arteriographic studies should be performed to confirm the diagnosis in those patients who are candidates for abdominal vascular surgery. The only definitive treatment is vascular surgery or balloon angioplasty to remove the thrombus or the construction of bypass arterial grafts to the ischemic bowel.

A number of systemic conditions are associated with *vasculitis* of the large and small arteries supplying the intestine. Most often these disorders can be recognized by the associated extraintestinal manifestations, as in polyarteritis nodosa, lupus erythematosus, dermatomyositis, Henoch-Schonlein purpura (allergic vasculitis), and rheumatoid vasculitis. When larger arteries are involved, as in polyarteritis nodosa, the picture of acute intestinal infarction is similar to that of embolic or atherosclerotic vascular occlusion. Often the involvement of smaller vessels leads to areas of intramural hemorrhage and edema resulting in abdominal pain, variable degrees of intestinal obstruction, and bleeding. Barium enema may show "thumbprinting" and "spiculation" due to localized edema, hemorrhage, and ulceration (Fig. 289-1). In many instances, treatment of the underlying disorder may lead to regression of symptoms. If signs of an acute abdomen develop, surgical exploration is usually indicated.

Intramural small-intestinal hemorrhage may occur with vasculitis, trauma, or impaired coagulation, especially in patients receiving anticoagulants. The clinical and radiologic features resemble those seen with vasculitis and local mucosal hemorrhage.

ISCHEMIC COLITIS

Ischemia of the colon most often affects the elderly because of their greater frequency of vascular disease. Ischemic colitis is almost always nonocclusive. Shunting of blood away from the mucosa may contribute to this condition, but the mechanism of ischemia is not known.

The clinical picture depends on the degree of ischemia and its rate of development. In *acute fulminant ischemic colitis*, the major manifestations are severe lower abdominal pain, rectal bleeding, and hypotension. Dilation of the colon and physical signs of peritonitis are seen in severe cases. Abdominal films may reveal thumbprinting from submucosal hemorrhage and edema (Fig. 289-1). Barium enema is hazardous in the acute situation because of the risk of perforation. Sigmoidoscopy or colonoscopy may detect ulcerations, friability, and bulging folds from submucosal hemorrhage. Angiography is not helpful in the management of patients with presumed ischemic colitis because a remediable occlusive lesion is very rarely found. Surgical resection may be required in some patients with fulminant ischemic colitis to remove gangrenous bowel; others with lesser degrees of ischemia may respond to conservative medical management.

Subacute ischemic colitis is the most common clinical variant of ischemic colonic disease. It produces lesser degrees of pain and bleeding, often occurring over several days or weeks. The left colon may be involved, but the rectum is usually spared because of the collateral blood supply, a feature distinguishing it from acute ulcerative colitis. Barium enema reveals edema, cobblestoning, thumbprinting, and occasionally superficial ulceration. Angiography is not indicated because almost all cases are nonocclusive. Occasionally, *stricture formation* may follow a bout of ischemic colitis or may present de novo without a history of antecedent pain or bloody diarrhea. Most cases of nonocclusive ischemic colitis resolve in 2 to 4 weeks and do not recur. Surgery is not required except for obstruction secondary to postischemic stricture.

ANGIODYSPLASIA OF THE COLON

These are vascular ectasias or arteriovenous malformations (AVMs) that occur in the right colon of many older individuals and may cause bleeding (Chap. 44). Angiodysplasia is a degenerative lesion consisting of dilated, distorted, thin-walled vessels lined by vascular endothelium. It may result from partial obstruction of the submucosal venous plexus by the tension generated in the cecal wall during muscular contraction. Grossly, angiodysplasias look similar to spider angiomas of the skin and on colonoscopy appear as star-shaped branching vessels in the submucosa measuring from 2 mm to 1 cm in diameter. The lesions are usually multiple and are found primarily in the cecum and ascending colon, but in some patients they may be distributed from the stomach to rectum.

Cecal angiodysplasia is important because of the likelihood of bleeding, either massively or chronically. In patients over age 60, ~1/4 of colonic bleeding episodes are secondary to angiodysplasia. The diagnosis is easiest to establish by colonoscopy, which allows treatment by laser photocoagulation, electrocautery, or injection with sclerosant. Some patients with massive uncontrolled bleeding or multiple sites of angiodysplasia may require right hemicolectomy. Angiodysplasias may also respond to

chronic estrogen-progesterone therapy.

ANORECTAL PROBLEMS

HEMORRHOIDS

The internal hemorrhoidal plexus of veins is located in the submucosal space above the valves of Morgagni. The anal canal separates it from the external hemorrhoidal venous plexus, but the two spaces communicate under the anal canal, the submucosa of which is attached to underlying tissue to form the interhemorrhoidal depression. Whenever the internal hemorrhoidal plexus is enlarged, the associated supporting tissue mass is increased, and the resultant venous swelling is called an *internal hemorrhoid*. When veins in the external hemorrhoidal plexus become enlarged or thrombosed, the resultant bluish mass is called an *external hemorrhoid*.

Both types of hemorrhoids are very common and are associated with increased hydrostatic pressure in the portal venous system, such as during pregnancy, straining at stool, or with cirrhosis. When internal hemorrhoids enlarge, pain is not a usual feature until the situation is complicated by thrombosis, infection, or erosion of the overlying mucosal surface. Most persons complain of bright red blood on the toilet tissue or coating the stool, with a feeling of vague anal discomfort. The discomfort is increased when the hemorrhoid enlarges or prolapses through the anus; prolapse is often accompanied by edema and sphincteric spasm. If not treated, prolapse usually becomes chronic as the muscularis stays stretched, and the patient complains of constant soiling of underclothing with very little pain. Prolapsed hemorrhoids may become thrombosed; the overlying mucous membrane may bleed profusely from the trauma of defecation.

Because they lie under the skin, external hemorrhoids are quite often painful, particularly if there is a sudden increase in their mass. These episodes result in a tender blue swelling at the anal verge due to thrombosis of a vein in the external plexus and need not be associated with enlargement of the internal veins. Since the thrombus usually lies at the level of the sphincteric muscles, anal spasm often occurs.

The diagnosis of internal and external hemorrhoids is made by inspection, digital examination, and direct vision through the anoscope and proctoscope. Since such lesions are very common, they must not be regarded as the cause of rectal bleeding or iron deficiency anemia until a thorough investigation has been made of the more proximal gastrointestinal tract. Acute blood loss can occasionally be attributed to internal hemorrhoids. Chronic anemia or occult blood in the stool in the presence of large but not definitely bleeding hemorrhoids requires a search for a polyp, cancer, or ulcer.

TREATMENT

Most hemorrhoids respond to conservative therapy such as sitz baths or other forms of moist heat, suppositories, stool softeners, and bed rest. Internal hemorrhoids that remain permanently prolapsed are best treated surgically; milder degrees of prolapse or enlargement with pruritus ani or intermittent bleeding can be handled successfully by banding or injection of sclerosing solutions. External hemorrhoids that become acutely

thrombosed are treated by incision, extraction of the clot, and compression of the incised area following clot removal. No surgical procedure should be carried out in the presence of acute inflammation of the anus, ulcerative proctitis, or ulcerative colitis. Proctoscopy or colonoscopy should always be performed before a patient undergoes hemorrhoidectomy.

ANAL INFLAMMATION

Perianal inflammatory lesions may be primary or may be associated with inflammatory bowel disease or diverticular disease. *Anal fissures* are superficial erosions of the anal canal which usually heal rapidly with conservative therapy. *Anal ulcers* are more chronic and deep and give symptoms largely as the result of painful spasm of the external anal sphincter during and after defecation. Bleeding may occur with either fissure or ulcer; healing of the ulcer is often associated with a hypertrophied anal papilla and some degrees of anal contracture. The spasm associated with chronic anal fissure/ulcer can be managed with oral nifedipine or local botulinum toxin. *Fistula in ano*, a tract leading from the rectal lumen to the perianal skin, usually results from local crypt abscesses. The fistula is a chronically inflamed canal made up of fibrous tissue surrounding granulation tissue, the lumen of which may be difficult to demonstrate. *Perirectal abscesses* often represent the tracking down into the anal area of purulent material escaping from the rectosigmoid; diverticulitis, Crohn's disease, ulcerative colitis, or previous surgery may be the underlying cause. Fistulas between the rectum and vagina or the rectum and bladder represent serious complications of granulomatous, septic, or malignant disorders and require the patient to be hospitalized for definitive diagnostic and therapeutic procedures.

PERITONEAL AND MESENTERIC DISEASES

ACUTE PERITONITIS

Peritonitis is a localized or generalized inflammatory process of the peritoneum that may appear in both acute and chronic forms. In the acute form the motor activity of the intestine is decreased, and the intestinal lumen becomes distended with gas and fluid. Fluid accumulates as a result of failure to reabsorb the 7 or 8 L normally secreted daily into the lumen and absorbed from the distal small bowel and colon. Because of accumulation of fluid in the peritoneal cavity as well as decreased oral intake, rapid depletion of the plasma volume with impaired cardiac and renal function may occur.

Etiology *Bacterial peritonitis* may be due to entry of bacteria into the peritoneal cavity from a perforation in the gastrointestinal tract or from an external penetrating wound. *Chemical peritonitis* results from spillage of pancreatic enzymes, gastric acid, or bile as a result of injury or perforation of the intestine or biliary tract. *Sterile peritonitis* occurs in patients with systemic lupus erythematosus, porphyria, and familial Mediterranean fever (FMF) during disease attacks.

The most common causes of bacterial peritonitis are appendicitis; perforations associated with diverticulitis; peptic ulcer; gangrenous gallbladder; and gangrenous obstruction of the small bowel from adhesive bands, incarcerated hernia, or volvulus. Any lesion leading to the escape of intestinal bacteria may be a source, including a

perforating carcinoma, foreign body, and ulcerative colitis. The peritoneal cavity is remarkably resistant to contamination, and unless continuing contamination occurs, the peritonitis remains localized. Patients with alcoholic cirrhosis and ascites have an increased susceptibility to *spontaneous bacterial peritonitis*, usually from enteric pathogens. This complication occurs in the absence of recognizable perforation of a viscus and may be due to leakage of bacteria through the intestinal wall (Chap. 299).

Clinical Features The cardinal manifestations of peritonitis are acute abdominal pain and tenderness. The location of the pain and tenderness depends on the underlying cause and whether the inflammation is localized or generalized. In *localized peritonitis*, as seen in uncomplicated appendicitis or diverticulitis, the physical findings are limited to the area of inflammation. With widespread peritoneal inflammation there is *generalized peritonitis* with diffuse abdominal tenderness and rebound. Rigidity of the abdominal wall is a common finding in peritonitis and may be localized or generalized.

Peristalsis may be present initially but usually disappears as the illness progresses and bowel sounds disappear. Hypotension, tachycardia, oliguria, and leukocytosis with cell counts>20,000/uL, are common, especially in generalized peritonitis. Plain abdominal films may reveal dilation of the large and small bowel with edema of the small-bowel wall, as evidenced by the distance between adjacent loops of gas-filled small intestine. Diagnostic paracentesis is sometimes valuable in determining the nature of the exudate as well as whether bacteria can be demonstrated or cultured.

GONOCOCCAL PERITONITIS

This usually involves an extension of gonococcal infection from a primary focus in the female reproductive tract. The signs of inflammation usually are limited to the pelvis, but there may be findings of a mild generalized peritonitis. Occasionally, the patient has right upper quadrant pain and tenderness caused by gonococcal perihepatitis involving the liver capsule and adjacent peritoneum (Fitz-Hugh-Curtis syndrome) (Chap. 147).

STARCH PERITONITIS

An acute granulomatous peritonitis can develop in some patients as a foreign-body reaction to cornstarch used to powder surgical gloves. The clinical picture is that of acute abdominal pain and fever 10 to 30 days after an abdominal operation. The diagnosis can be made by paracentesis and demonstration of starch granules in monocytes. However, most patients are reexplored because of the fear of abscess or bacterial peritonitis, with the finding of foreign-body granuloma studding the peritoneum.

PSEUDOMYXOMA PERITONEI

This is a rare condition resulting from rupture of a mucocele of the appendix, a mucinous ovarian cyst, or mucin-secreting intestinal or ovarian adenocarcinoma. The abdomen becomes filled with masses of jelly-like mucus. Occasional patients are cured with removal of the mucocele or the ovarian cyst and most of the myxomatous material. In other cases, however, the mucoid material recurs, leading to progressive wasting and eventual death. Colloid carcinoma arising from the stomach or colon with peritoneal implants may resemble pseudomyxoma at laparotomy. The course of this type of highly

malignant tumor is one of rapid cachexia and early death. The diagnosis usually can be made by the appearance of many highly malignant cells in the peritoneal implants.

PNEUMATOSIS CYSTOIDES INTESTINALIS

This is a condition in which multiple gas-filled blebs or cysts accumulate in the intestinal wall beneath the serosal surface of the bowel. The exact source of the gas has not been explained satisfactorily. In some instances, this disease is associated with specific ulceration of the intestinal mucosa, in particular peptic ulcer with outlet obstruction. Cysts in the wall of the small bowel are seen as an occasional complication of mesenteric vascular occlusion. In the large bowel, these cysts are usually benign, may be seen with a variety of other disorders, and usually disappear over time.

Physical findings are not specific and the diagnosis is made either by x-ray or at laparotomy. Occasionally, the subserosal cysts may rupture, resulting in pneumoperitoneum.

CHYLOUS ASCITES

SeeChap. 46.

MESENTERIC LIPODYSTROPHY

This is a rare disorder usually affecting middle-aged women and characterized pathologically by infiltration of the mesentery with lipid-laden macrophages and fibrous tissue. These patients present with ill-defined abdominal pain and occasionally an abdominal mass. The diagnosis is made at laparotomy by demonstration of thick fibrofatty masses at the root of the mesentery with retraction and distortion of the bowel loops.

FAMILIAL MEDITERRANEAN FEVER

Familial Mediterranean fever (FMF, familial paroxysmal polyserositis) is an inherited disorder, characterized by recurrent episodes of fever, peritonitis, and/or pleuritis. Arthritis, skin lesions, and amyloidosis are seen in some patients.

FMFoccurs predominantly in patients of non-Ashkenazi (Sephardic) Jewish, Armenian, and Arabic ancestry. However, the disease has been seen in patients of Italian, Ashkenazi Jewish, and Anglo-Saxon descent as well as others.

ETIOLOGY

FMFis an autosomal recessive trait characterized by mutations in the *MEFV* gene located on 16p. The gene encodes a 781-amino acid protein called *pyrin* expressed in cells of the myeloid lineage. The gene is in the *RoRet* family, and the product appears to function as a transcription factor based on the presence of a nuclear localization signal, a zinc finger, and a coiled-coil domain. Its expression in granulocytes is increased by proinflammatory cytokines and reduced by anti-inflammatory cytokines. Mutations associated with FMF cluster in exon 10. Different mutations appear to be associated

with distinct disease manifestations; replacement of methionine 694 by valine is common in patients who have amyloidosis as a feature of the disease. Valine 726 replacement by alanine is rarely associated with amyloidosis. Other as yet unknown genes may modify the phenotype or be responsible for FMF in non-Mediterranean populations.

PATHOLOGY

Despite the striking clinical manifestations during an acute attack ofFMF, no specific pathologic alterations have been found. At laparotomy there is acute peritoneal inflammation with an exudate that contains a predominance of polymorphonuclear leukocytes. A disproportionately large number of male patients develop gallbladder disease with and without cholelithiasis. Pleural and joint inflammation are also nonspecific.

In the amyloidosis that accompaniesFMF, amyloid is deposited in the intima and media of the arterioles, the subendothelial region of venules, the glomeruli, and the spleen. Aside from their vessels, the heart and liver are uninvolved.

MANIFESTATIONS

The symptoms ofFMFoften begin between the ages of 5 and 15, although attacks sometimes commence during infancy and onset has occurred as late as age 50. The duration and frequency of attacks vary greatly in the same patient, and their occurrence follows no set pattern. The usual acute episode lasts 1 to 2 days, but some may be prolonged for 7 to 10 days. The attacks range in frequency from twice weekly to once a year, but 2 to 4 weeks is the most common interval. Spontaneous remissions lasting years have been seen. The severity and frequency of the attacks decrease with age or with development of amyloidosis.

Fever Fever is a cardinal manifestation and is present during most attacks. Rarely, fever may be present without serositis. The temperature may be preceded by a chill and will peak in 12 to 24 h. Defervescence is often accompanied by diaphoresis. The fever ranges from 38.5° to 40°C but is quite variable.

Abdominal Pain Abdominal pain occurs in >95% of patients and may vary in severity in the same patient. Minor premonitory discomfort may precede an acute episode by 24 to 48 h. The pain usually starts in one quadrant and then spreads to involve the whole abdomen. The initial site is usually very tender. Tenderness may remain localized with referred pain in other areas, and may radiate to the back. Diaphragmatic irritation may lead to splinting of the chest and pain in one or both shoulders. Nausea and vomiting sometimes occur. The abdomen is usually distended and may become rigid, with decreased or absent bowel sounds. On x-ray the wall of the small intestine may appear edematous, transit of barium is slowed, and fluid levels may be seen. An abdominal operation may precipitate an acute attack ofFMF, which may be confused with other postoperative complications.

Chest Pain Most patients with abdominal attacks have referred chest pain at one time or another, and 75% also develop acute pleuritic pain with or without abdominal

symptoms. In 30%, the attacks of pleuritis precede the onset of abdominal attacks by varying periods of time, and a small number of patients never develop abdominal attacks. Chest pain is usually unilateral and is associated with diminished breath sounds, a friction rub, or a transient pleural effusion.

Joint Pain In Israel, 75% of patients report at least one episode of acute arthritis. Arthritis can be distinct from abdominal or pleural attacks, can be acute or, rarely, chronic, and may involve one or several joints. Effusions are common, with large joints most frequently involved. Radiologic findings are nonspecific. Despite careful search, frank arthritis rarely has been seen in the United States. Some patients have a history of rheumatic fever-like illness in childhood, but in a large series of patients, including 30 from the Middle East, acute arthritis was not observed. Mild arthralgia is common during acute attacks but is nonspecific.

Skin Manifestations Skin involvement occurs in one-third of patients. These lesions consist of painful, erythematous areas of swelling from 5 to 20 cm in diameter, usually located on the lower legs, the medial malleolus, or the dorsum of the foot. They may occur without abdominal or pleural pain and subside within 24 to 48 h.

Other Signs and Symptoms Involvement of other serosal membranes has been reported, but pericarditis and meningitis are rare. Hematuria, splenomegaly, and small white dots called *colloid bodies* in the ocular fundus are findings of questionable significance. Rarely, migraine-like headaches accompany acute abdominal attacks, and some patients have become somewhat irrational or show extreme emotional lability during attacks. Whether these are primary manifestations ofFMF or secondary effects of pain and fever is unclear.

Complications Depression and lack of motivation are common, and patients withFMFrequire considerable support. A striking number of patients have developed gallbladder disease.

Amyloidosis has been reported in Israel, North Africa, and elsewhere in the Middle East, but its occurrence is rare in the United States. These findings are even more striking because there are probably as many knownFMFpatients in the United States as in Israel. Thus, environmental or nutritional, as well as genetic, factors may play a role in the development of amyloidosis in FMF.

LABORATORY FINDINGS

Polymorphonuclear leukocytosis ranging from 10,000 to 30,000 cells/uL is almost invariable during acute attacks. The erythrocyte sedimentation rate is elevated during attacks but returns to normal between attacks. Plasma fibrinogen, serum haptoglobin, ceruloplasmin, and C-reactive protein increase during the episodes. Plasma lipids are normal, and no consistent abnormalities of hepatic or renal function are seen. When amyloidosis is present, laboratory findings are typical of a nephrotic syndrome followed by renal insufficiency.

DIAGNOSIS

When the typical acute attacks ofFMFoccur in an individual of appropriate ethnic background with a family history of FMF, the diagnosis is easy. When a patient is seen for the first time, a variety of other febrile illnesses must be excluded, such as acute appendicitis, pancreatitis, porphyria, cholecystitis, intestinal obstruction, and other major abdominal catastrophes.

Some inherited hyperlipidemias may mimic the clinical picture ofFMF, but lipid analysis will eliminate them from consideration. The FMF patient is not immune to other diseases, and when an attack differs from the usual pattern or is more prolonged, consideration should be given to other diagnostic possibilities. The pleural form of the disease is sometimes difficult to differentiate from acute pulmonary infection or infarction, but the rapid disappearance of signs and symptoms resolves the problem. The erythema is sometimes difficult to differentiate from superficial thrombophlebitis or cellulitis.

The most difficult diagnostic problem inFMF is the patient who presents with fever alone. In this situation an extensive diagnostic workup for fever of unknown origin may be required. Fortunately, such patients are rare, and all eventually develop serosal involvement. Until specific diagnostic tests for FMF are available, patients with recurrent fever but without signs of inflammation of one of the serosal membranes should not be categorized as having FMF.

PROGNOSIS

Despite the severity of the symptoms during some acute attacks, most patients are remarkably free of debilitation between attacks and are able to lead fairly normal lives. The greatest hazard to patients is prolonged periods of hospitalization due to erroneous diagnoses or failure to understand the disease. In the United States, the prognosis of patients withFMF does not seem to be different from that of patients with other chronic nonfatal illnesses. Death usually results from causes unrelated to the underlying disease.

In the past, ~25% ofFMFpatients in Israel developed amyloidosis, and this complication usually led to death. However, the widespread use of colchicine has resulted in a dramatic decrease in the incidence of amyloidosis.

TREATMENT

During the past 25 years, the outlook of patients withFMF has been altered dramatically. Chronic administration of colchicine greatly reduces the number of acute attacks of FMF. It is recommended that 0.6 mg colchicine be taken by mouth three times a day. If patients develop gastrointestinal side effects with this dose, it should be reduced to 0.6 mg taken twice a day. Although an occasional patient will respond to 0.6 mg taken only once a day, this amount is less likely to be beneficial. Most FMF patients will respond favorably to colchicine prophylaxis.

(Bibliography omitted in Palm version)