Nutshell Series

Health Care Law and Ethics

in a nutshell®

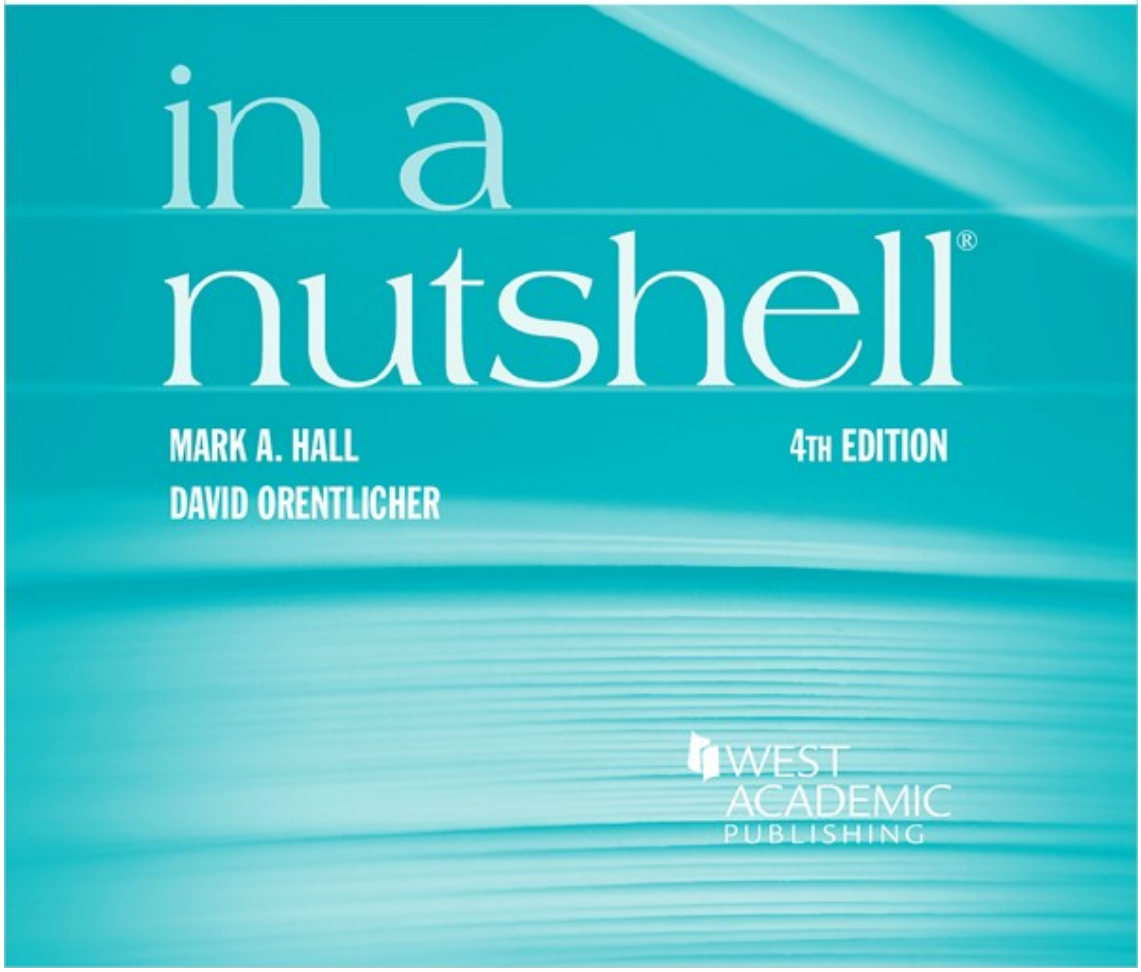
MARK A. HALL
DAVID ORENTLICHER

4TH EDITION

WEST ACADEMIC PUBLISHING

Nutshell Series

Health Care Law and Ethics



MARK A. HALL
DAVID ORENTLICHER

4TH EDITION

WEST ACADEMIC PUBLISHING

WEST ACADEMIC PUBLISHING'S LAW SCHOOL ADVISORY BOARD

HEALTH CARE LAW AND ETHICS

IN A NUTSHELL®

FOURTH EDITION

MARK A. HALL
Fred and Elizabeth Turnage
Professor of Law and Public Health
Wake Forest University

DAVID ORENTLICHER
Cobeaga Law Firm Professor of Law
UNLV William S. Boyd School of Law

WEST ACADEMIC PUBLISHING

444 Cedar Street, Suite 700
St. Paul, MN 55101
1-877-888-1330

West, West Academic Publishing, and West Academic are trademarks of West Publishing Corporation, used under license.
Printed in the United States of America
ISBN: 978-1-68467-642-2

In memory of friend and colleague,
Dan Strouse, formerly Professor of Law at Arizona State University

PREFACE

Scholars debate whether health care law is a coherent intellectual field. Some call it a "chaotic, dysfunctional patchwork" of doctrine, whereas others believe the field coheres around the essential features of medicine and how they systematically shape and change generic law. Despite this unresolved debate, there is a good deal of consistency about the content of health care law courses taught not only in law schools, but also schools of medicine, public health, and health care administration. Typically, these courses divide into three or four distinct subject areas. Historically, "law and medicine" courses were dominated by the subject of medical malpractice and public health law. These subjects remain important, and are the topic of other books in the Nutshell series. But they no longer are the dominant focus of the field. Modern health care law courses also include two areas that developed rapidly and extensively over the past generation (or two): the corporate, regulatory, and financial structure of health care delivery, and bioethics. It is these two areas that this book addresses. The increasing importance of both subjects is due in part to the incredible advances in medical technology in the past half century. These advances have had two subsidiary effects. One is a phenomenal increase in the cost of health care. Because doctors can now routinely do so much more for their patients, the cost of medicine has skyrocketed, fueled in part by the development of dramatic new technologies such as organ transplants, kidney dialysis, and open heart surgery. These same medical advances also generate many additional ethical problems, for we did not confront the question of a whether a patient's life should be extended by

kidney dialysis or open heart surgery before those technologies became available.

Part I of this book contains foundational chapters that are relevant to any course in health care law. It capsules the legal and policy issues of health care funding, access, and reform, and it explores the legal structure and content of the doctor-patient relationship. The dramatic growth of health care spending and efforts to contain spending lead to both innovations and tensions in the business of health care. Part II addresses the legal issues that result from these structural and economic forces. Part III moves from these "macro" issues to the ethical dilemmas that arise in individual patient decisions, and covers most of the topics addressed by law school courses in bioethics.

Most schools teach the financial and structural issues of health care delivery in a separate course from the bioethical puzzles, often through different instructors. Other instructors are more generalists who combine elements of both these areas with the traditional medical malpractice topics into a successful overview course. To maximize our respective talents, we have divided primary authorship responsibilities along these same lines (Hall—Chapters 1–5; Orentlicher—Chapters 6–8). Also, we owe a large intellectual debt to our colleague, Ira Mark Ellman, who had primarily responsibility for the bioethics chapters in prior editions. Working together on this book has persuaded us, however, that this tripartite field of law has many common themes that can and should be integrated into a more coherent whole. To that end this book is a small step.

MARK A. HALL
DAVID ORENTLICHER

January 2020

A GLOSSARY OF TERMS AND ACRONYMS

The literature of health care law, and especially health care financing, is filled with acronyms and specialized terminology. We have avoided these where possible, but they are sometimes necessary, for many concepts and agencies are better known by their acronym or specialized phrase than by their full name. To ease the burden on the reader new to the subject, the following list gathers the major acronyms and terms of art used in this book primarily in the first five chapters:

ACA: Affordable Care Act (a.k.a. "Obamacare")—A comprehensive reform of the individual health insurance market, a substantial expansion of Medicaid, and a significant body of additional requirements for group insurance.

ACO: Accountable Care Organization—A newly-emerging (still ill-defined) construct in which physicians and/or hospitals align in order to receive insurance payment on a bundled rather than fee-for-service basis, and to be jointly responsible in some fashion for the quality of care.

AMA: American Medical Association.

AHA: American Hospital Association.

Capitation: A payment method for physician or hospital services that pays a fixed amount per person, regardless of how much service they need.

CMS: Centers for Medicare and Medicaid Services, formerly known as HCFA, the Health Care Financing Administration, is the

agency within DHHS with direct responsibility over Medicare and Medicaid.

CON: Certificate of Need

DHHS: U.S. Department of Health & Human Services

DRGs: Diagnosis-Related Groups—The reimbursement method recently adopted by Medicare, which pays hospitals a single, preset amount for each patient admitted according to the patient's diagnosis, age, and condition.

ERISA: Employee Retirement Income Security Act of 1974—A federal statute that primarily regulates private pension plans, but also encompasses other employee benefits such as health insurance. It has tremendous importance for the regulation of health care delivery because of a sweeping provision that preempts many traditional sources of state law.

HMO: Health Maintenance Organization—A health care organization that combines insurance and treatment functions in the same entity by providing all needed care for a lump sum annual payment.

IDS: Integrated Delivery System—Any one of a number of innovative techniques for combining hospital, physician, and insurance components into a single structure or organization.

IPA: Individual Practice Association—A form of HMO that provides care through contracting physicians who maintain independent practices in their individual offices. This structure contrasts with group HMO models in which physician owners or employees operate out of a clinic-based setting. The terms is also used to refer to a physician group that contracts with insurers, usually on a capitation basis.

JCAHO: The Joint Commission on Accreditation of Healthcare Organizations—A private credentialing organization with enormous influence on hospital structure and functioning.

Managed Care: Any system of health service payment or delivery arrangements in which the health plan or provider attempts to control or coordinate health service use to contain health expenditures, improve quality, or both. Arrangements often involve a defined delivery system of providers having some form of contractual relationship with the plan.

Managed Competition: A system for choosing and paying for health insurance in which subscribers have a choice among a variety of different plans, but they must pay the most or all of the difference in price between the plan they choose and the least expensive one. Insurers in turn are required to accept all applicants and to not charge sick people more. In theory, this will force insurers to compete by managing the cost of care rather than on their ability to select only the best risks.

PHO: Physician-Hospital Organization—Any one of a number of different types of joint ventures between hospitals and physicians, usually intended as a contracting vehicle for forming a managed care network.

PPO: Preferred Provider Organization—A hybrid between traditional indemnity insurance and HMOs, in which subscribers still have a choice of physicians and providers are still paid fee-for-service, but subscribers are encouraged through financial incentives to use a select group of providers in a network, and providers agree to accept discounted payment. Similar to a point-of-service (POS) plan, which has the same elements but is licensed as a type of HMO that allows subscribers to go outside the network by paying higher deductibles and copayments.

PPS: Prospective Payment System—Another description of Medicare DRGs, one that emphasizes the fixed, preset nature of payment.

RBRVS: Resource-Based Relative Value Scale—A fee schedule used by Medicare to pay physicians, which is based on the skill and effort required for each service rather than on historical billing patterns.

OUTLINE

TABLE OF CASES

References are to Pages

HEALTH CARE LAW AND ETHICS

IN A NUTSHELL®

FOURTH EDITION

PART I

THE FUNDAMENTALS OF HEALTH CARE DELIVERY AND FINANCE

These are exciting but challenging times in which to study health care law. Traditionally, the exclusive focus of health care policy was on advancing the state of medical science. Over the past generation, however, two new concerns have begun to dominate the health care policy agenda: the cost of and access to medical treatment. Limiting incessant inflation in health care spending for the bulk of the population covered by insurance and finding ways to afford health care coverage for the medically indigent are issues of paramount importance in the decisionmaking of law makers, medical practitioners, and health care institutions. These issues were brought to a head a decade ago in Congress' enactment of the Patient Protection and Affordable Care Act of 2010 (Affordable Care Act or ACA). Over that decade, the ACA survived repeated challenges both in Congress and in the courts, yet pressing issues of medical costs and access remain acute in health care public policy, law, and ethics.

The fundamental reorientation in perspective caused by the shift in focus from advances to restraints demands a thorough rethinking of traditional legal doctrine from the ground up. Legal precedents rooted in the expansionist medical care system of the post-war era sometimes no longer make

sense in an era when the controlling public policy is to limit or allocate governmental and private health care spending. For example, perhaps malpractice law should take account of the economic costs as well as the medical risks of treatment decisions. Or perhaps antitrust, tax and corporate law should be take more account of health policy concerns in determining how medical institutions are structured and operated. In the field of bioethics, perhaps the law still has not found the right balance among patient autonomy, professional prerogative, and institutional integrity.

So that the reader may acquire a deeper understanding of these new legal challenges, Chapter 1 explores the policy and economic environment that shape the agenda for contemporary health care law and ethics. Chapter 2 then looks at the areas of legal doctrine that directly affect a patient's right to receive health care, and the structure and content of the doctor-patient relationship.

CHAPTER 1

HEALTH INSURANCE COVERAGE AND REGULATORY REFORM

A. THE CRISIS IN HEALTH CARE SPENDING AND COVERAGE

1. THE DIMENSIONS OF THE CRISIS

a. The Spending Crisis

An astounding rate of unrelenting inflation has gripped the American health care system over the past half century. Medical spending's portion of the overall economy has nearly quadrupled since 1960. We now spend over 3 trillion dollars annually on health care, which is about $10,000 per person, and 18 percent of the gross national product—larger than any other economic sector. As startling as the present day facts are, these trends are even more disturbing considering the prolongation of life spans and the vast increase in elderly population that will occur with the aging of the baby boom generation.

These increases in spending might be celebrated as a great American success story if we believed they produced commensurable benefits. After all, nothing is more treasured than our health; why shouldn't we spend as much as possible to enhance it? There are three responses to this observation. First, all spending is inherently a trade-off because we can always spend on something else instead. While our health is surely an important priority, it simply is not true that people always treat health as more

important than competing values. We see this every day, not only when we smoke, drink, or eat excessively, but also when we drive to work or we read a book rather than exercise. Second, even if we focus on our health as the overriding value, we must realize that there are other ways to promote health than through medical care. In the words of one prominent health policy analyst, "once a reasonable minimum level of care is provided, factors other than medical care—diet, lifestyle, heredity, environment—appear to have much more effect on health and longevity than does more or less medical care. Above a reasonable minimum, the availability of more medical care resources appears to have little or no effect on many indicators of health status." A. Enthoven, Health Plan xvi (1980). In other words, we are likely to produce greater health bang for the buck by spending less on medicine and more on nutrition, shelter, the environment, education to improve lifestyles and other preventive measures.

The idea that American medicine is an unqualified success story is also undermined by international comparisons. We spend substantially more per capita and devote much more of our GDP to health care than any other peer nation. Our proportionate spending is 50-to-100 percent higher than most industrialized countries. At the same time, the U.S. ranks near the bottom in major indicators of health status such as life expectancy at birth and infant mortality. To some extent, this is due to social factors that create health problems. By other indicators American medicine is the best in the world. For instance, the chance of dying once a person has a major disease is

substantially lower in America than in many other countries. Still, it appears that Americans are not getting the same bang for their medical buck as are other countries. It has been documented that much of what we spend on medical care is simply waste, yielding no benefit of any kind. Examples of unnecessary care proliferate throughout American medicine. According to some accounts, a quarter to a third of health care spending may be completely unnecessary, and another substantial increment may be of questionable or only marginal value.

b. The Coverage Crisis

Medical spending is not the only health care crisis gripping our country. We are also facing a crisis of epidemic proportions in health insurance coverage. At the same time that we have devoted massive resources to the health care of most of our populace, we have largely ignored the health care needs of a significant segment of our society. The ranks of the medically indigent increased dramatically to the point that, in 2018, an estimated 50 million adults (about 25 percent of those under age 65) were uninsured at some point during the year. (Only about half that number were uninsured the *entire* year, however.)

This occurred because many employers do not offer their workers private health insurance as a job benefit, especially smaller employers and those that hire seasonal and temporary workers. Workers without insurance can buy individual coverage, but that is more expensive than group coverage, and

these workers often have low-wage jobs or are young and healthy and so do not perceive the need for health insurance.

The gaps in private insurance are partially filled by two major public programs, Medicaid and Medicare, but Medicaid covers only those who are poor and Medicare covers the elderly or disabled. Medicare is reasonably comprehensive. Almost everyone who is over 65 and retired is eligible, and those who aren't can buy in for a reasonable fee. But Medicaid historically fell far short of its objective. Prior to the Affordable Care Act (ACA), Medicaid covered less than half of people falling below poverty (as defined by federal guidelines). This was due to two factors. First, Medicaid is a joint program between state and federal governments, and individual states set different income guidelines. Second, federal rules historically restricted Medicaid eligibility to specified categories of poverty that attempt to define who are the "deserving poor." Simple poverty did not suffice in most states; one also had to be elderly, disabled, or the parent of a dependent child. These are the traditional categories covered by welfare.

Starting in 2014, Affordable Care Act (a.k.a. "Obamacare") required states to expand Medicaid to cover all legal residents up to 138 percent of the federal poverty level and paid for 90 percent of the costs of this expansion. The Supreme Court, however, ruled that requiring states to pay 10 percent of expansion costs violates states' rights under the 10th Amendment. NFIB v. Sibelius, 567 U.S. 519 (2012).

Accordingly, about a third of states, mostly in the south and midwest, have refused to expand Medicaid, for a mix of ideological and political reasons.

Other components of the ACA provide sliding-scale subsidies for people up to 400 percent of poverty, to purchase private insurance

(if they do not have insurance through their work). This subsidized private insurance is available through on-line insurance "exchanges" in each state, created by the ACA. But, because the ACA assumed all states would expand Medicaid to cover all legal residents below poverty, the subsidized exchanges are not available to people who are below the federal poverty line. This results in a somewhat inexplicable coverage gap in non-expanding states, such that insurance is readily available to people just above poverty but not at all to those below poverty, unless they fit one of Medicaid's traditional categories of need.

As a consequence, almost 30 million people will remain uninsured through the year. These are people in the Medicaid coverage gap, undocumented immigrants ("illegal aliens"), people who neglect to sign up for affordable coverage, and people whose insurance is still unaffordable because they do not qualify for subsidies or employer coverage. Thus, despite this comprehensive insurance reform, the U.S. remains the only industrialized country without universal access to care. The ACA's reforms improved insurance coverage substantially, but so far it has cut the number of uninsured people less than half, down to eight-nine percent of the population.

When uninsured people are ill, they must either pay out of pocket, go to publicly-funded or volunteer clinics, or seek charitable care from hospitals, usually in the emergency room.

Coverage is even more deficient for certain categories of care. Many well-insured workers have little or no coverage for long-term care in nursing homes—an issue of special concern as the ranks of the very old are beginning to escalate. Traditional health insurance,

including Medicare, covers nursing home and home health care only for short periods of time. Long-term care of this nature requires separate insurance, which most people do not think to purchase until they need it, which makes the price not worth it or too high to afford for most people. Some can afford to pay for care out of pocket, but the only fall-back for many people is Medicaid. A major portion of Medicaid funding goes to cover the costs of nursing home care for the middle class elderly. These people qualify by first "spending down" their available wealth on the first few years of nursing care and, once impoverished, then going onto Medicaid. For a time, careful estate planning could produce the same result by transferring assets to trusts or to family members, but a series of statutory amendments have virtually eliminated the ability to qualify without actually spending down to below the poverty line.

2. THE CAUSES OF THE SPENDING CRISIS

Although the Affordable Care Act made *insurance* much more affordable for many Americans, it did not

reduce the underlying costs of health *care*. To secure the support or avoid the opposition of key provider interest groups, the Obama administration refrained from imposing new controls on the price of medical care. Thus, the crisis in medical costs continues to make it difficult to extend comprehensive coverage to everyone. Therefore, the first task of health care policy is to understand the causes and possible solutions to the relentless inflation that has gripped health care spending over the past generation. After that, we will look at proposals for making access to health insurance more comprehensive.

a. The Structure of Traditional Insurance

Health policy analysts agree that the central cause of the uncontrolled increase in health care spending is the "complex of irrational economic incentives" inherent in our conventional health insurance system. A. Enthoven *supra* at 16. Some three-fourths of medical treatment is funded by third-party payers—either private insurance policies available primarily through employment or government programs available to the elderly and poor. Traditionally, health insurance has been structured on a piece-work basis known as "fee-for-service," whereby doctors, hospitals and other providers are paid a separate amount for each discrete item of service. This traditional form of reimbursement creates powerful cost-escalating incentives that affect each actor in the health care system.

b. The Incentives for Patients

The patient has virtually no incentive to economize under this system of insurance. Because a third party is footing most of the bill, patients are eager to receive (and have now come to expect and demand) all care that is of any conceivable benefit. In economic terminology, this tendency of insurance to induce disregard of costs is known as a form of "moral hazard" (analogous to the incentive that a well-insured warehouse owner might have to avoid taking precautions against fire). Although it is debatable whether well-insured patients may be less concerned about taking preventive measures, it is indisputable that traditional insurance induces them to order more treatment than they otherwise would when they do fall ill or suffer an accident.

One might expect insured patients to scrutinize medical expenditures in order to avoid future premium increases. This is one deterrent to careless driving under auto insurance, for example. The reason this cost-internalization safeguard does not work is the lack of "experience rating" in health insurance. Rather than individualizing premiums to the specific medical treatment record of particular patients (as auto insurers do for drivers), health insurers historically engaged in "community rating," a system that charges the same for everyone in the insurance pool. Blue Cross/Blue Shield, the largest network of health insurers, was notable in this regard. Other insurers, while individualizing premiums to some extent, still use large groups of employees as their basis for setting group rates. Even

to the extent that health insurance is experience rated, *patients* still are not likely to realize the financial impact of their treatment decisions since the great bulk of insurance is provided in the employment setting. Therefore, the employer is the one who picks up the tab for insurance premium increases. Moreover, the fact that employer-paid insurance premiums are not taxed as income to the employee creates a strong tax incentive for employees to prefer lower wages over being required to pay a portion of the premium themselves.

Another way to describe this set of incentives is through the economic concept of the "free rider" effect. Because the costs of treatment decisions are not internalized to the patient but instead are spread to others in the employee group or the community, individual patients are able to take a free ride when they spend excessively. This effect is most pronounced under government

insurance, where the costs of treatment are financed through the general tax base and therefore borne by society at large.

"[O]ne wonders what people thought *would* control cost in such a system. One of the main answers is deductibles and coinsurance. Make the patient pay the first $200 of each year's medical bills and 25 percent of the cost above that, and he will be cost-conscious and go to the doctor only when necessary. . . . This principle has been applied in most health insurance in the United States." A. Enthoven, Health Plan 32–33 (1980). This cost-sharing principle has not worked well, even when used, because most people base medical decisions on

what their doctors recommend, and they lack the knowledge and expertise to competently evaluate their doctors' treatment recommendations. It is a perverse market that places purchasing decisions in the hands of sellers rather than buyers, but this phenomenon of "supplier-induced demand" is what characterizes the traditional market for health care services. Thus, as a generalization, it is probably more accurate to view the health care system as driven by the decisions of physicians than by those of patients. For this reason, it becomes critical to understand the set of incentives that traditional reimbursement creates for health care providers.

c. The Incentives for Providers

To the extent that physicians control treatment decisions, the third-party character of insurance—the fact that someone else is footing the bill—results in a lack of cost consciousness in health care purchasing decisions, reinforcing the economically absurd "spare-no-expense" philosophy just described. But this is not the

entire story. Traditional insurance is riddled with other incentives that are equally or more perverse with respect to physician behavior. Primarily, the particular structure of the payments made under traditional insurance has a critical influence on medical practice. Fee-for-service reimbursement pays doctors and hospitals more the more they treat. This "piece work" payment method has three results: First, it exacerbates the "spare-no-expense" ethic described above to the extent that providers have an incentive to render not only all care that has any benefit, but also care that may be

of *no* benefit (or only of very uncertain benefit). Second, fee-for-service reimbursement deemphasizes preventive care, which is usually not as lucrative for providers. Third, this payment system encourages an excessive reliance on medical technology by paying for discrete procedures rather than for the time spent with patients and by encouraging the medicalization of social problems. Traditional insurance has thus been blamed for the current ills of overspecialization, excessive reliance on hospitalization, and triple-teaming cases.

This is not to say that doctors and hospitals act only out of economic motives. Indeed, providers may not even consciously consider the economic implications of their actions. Nevertheless, the economic environment created by third-party, fee-for-service reimbursement certainly facilitates and reinforces an inflationary practice style. It is impossible to expect doctors to be impervious to these fundamental economic forces. Indeed, it might be considered unethical for doctors to act in any other way. Ethicists maintain that a doctor's fiduciary duty to his patients requires him to act solely with the patient's interests in mind. If costs are of no consequence

to the patient, then they should not enter into the doctor's calculus. But if costs are of no consequence to the patient or the doctor, then who will insist upon efficiency and value in medical expenditures? This oversight role is left to third parties—employers and insurers (both private and government).

d. The Incentives for Insurers and Employers

In a system of third-party reimbursement funded by employers and the government, both they and the insurers have strong incentives to restrain costs. Yet, several dynamics within the health care system have made it difficult for third parties to impose meaningful discipline. First, the prevalence of the "free choice of provider" principle kept insurers from using competitive bidding to impose cost discipline on the medical profession or the hospital industry. Organized medicine campaigned heavily throughout the 20th century to give patients the freedom to select any doctor or hospital they desire while maintaining comprehensive insurance. Some commentators argue that this ostensible consumer choice ethic has been misused to achieve anticompetitive ends, because requiring insurers to reimburse most or all providers eliminates any opportunity for insurers to select relatively more efficient doctors or hospitals through a competitive bidding process.

Private and government insurers might attempt to regulate provider behavior by rigorous scrutiny of the need for medical services through the process of individual claims review. Such "utilization review" techniques face the obstacle of developing acceptable methods for identifying unnecessary care, a task made difficult by the fact that only a very small portion of medical care is of a life-or-death nature. The vast bulk of medical decisions are

focused on questions such as whether an extra diagnostic test or an extra day in the hospital is worthwhile, questions

15

to which there are no clear answers. At the margin, it is extremely difficult to say that a particular item of service has no benefit whatsoever. It is even more difficult to say when a procedure is not worth its cost if it has *some* benefit.

Such assessments are rendered impossible in part by the extent of uncertainty in medical knowledge. This uncertainty results in large part from the difficulty of conducting elaborate controlled studies to test the relative benefits of all the many options available to treat each medical condition. Without a firm, scientific basis on which to critique medical decisionmaking, insurers are forced as a practical matter to acquiesce in a system that allows the very persons who are receiving reimbursement to exercise the sole authority over certifying the legitimacy of the expenditures. Even if dispositive studies were available, physician oversight would be deterred by the intensely judgmental nature of medical practice. This notion is captured in the slogan that medical practice is more an art than a science, meaning that physicians' treatment decisions are guided more by soft, subjective reasoning processes than by rigorous, deductive logic. As a result, it is difficult for someone other than the patient's personal physician to dictate the details of treatment, even if a third party's judgment were equally valid from an objective perspective. The conventional consequence of these various factors was to cede to practicing physicians nearly complete authority over treatment decisions.

16

e. Traditional Medicare Reimbursement

The structure of traditional Medicare reimbursement for hospitals provides an excellent microcosm of the failings of conventional insurance. Medicare, which covers the elderly and disabled regardless of wealth or income, was enacted during the liberal euphoria of the 1960s. It has two basic parts: Part A (the focus of the following discussion), covers the services of hospitals and other facilities; Part B covers physician and outpatient services. Part D, enacted in 2006, covers prescription drugs, and Part C (also called Medicare Advantage) is an option that allows people to receive all three parts of their Medicare benefits through a private insurer such as an HMO.

The hospital industry and the medical profession lobbied against Medicare (and Medicaid) when these programs were in their planning stages because hospitals and doctors viewed large-scale government intrusion into medicine as a potential threat to their interests. Congress, concerned that the necessary level of cooperation would not be forthcoming, responded to this pressure by crafting Medicare with several elements favorable to providers.

Primarily, Medicare was modeled after traditional private insurance with fee-for-service payment and free choice of physician. For doctor services, Part B of Medicare pays 80 percent of "reasonable charges." Part A initially reimbursed hospital services 100 percent of the costs of treatment less a deductible and copayments, but this was changed in 1983, as explained more below, because of the obvious

inflationary incentives. Moreover, the particular measure of cost that was used provided ample allowance for rapid depreciation of

assets and even paid profit-making hospitals an amount reflecting a return on investment. Plus, the government decided not to undertake itself the task of administering reimbursement, instead contracting out these accounting details to "fiscal intermediaries," usually Blue Cross and Blue Shield, which are creatures of the hospital industry.

As a result of these various fundamental structural elements, Medicare expenditures quickly mushroomed far beyond initial projections. The facts of Sacred Heart Hospital v. United States, 616 F.2d 477, 479 (Ct.Cl.1980), a typical reimbursement dispute at one hospital, provide a representative illustration: "Prior to 1963, . . . the [respiratory therapy] department, as such, consisted of a non-certified therapist and a technician who were trained to administer oxygen. . . . [Since then], the department has grown to eight full time, board certified anesthesiologists, nine trained technicians, two therapists and two registered therapists."

f. Managed Care and the Structure of Health Insurance

Considering that the structure of health insurance lies at the root of so many of these economic and social problems, it makes sense to ask why this structure dominated for so many decades. Medical sociologist Paul Starr and other scholars have documented a sustained and enormously successful

campaign by organized medicine to suppress or co-opt alternative insurance models that would have been more threatening to its economic interests. *See, e.g.,* Paul Starr, The Social Transformation of American Medicine 226 (1982). Generally, "of the variety of insurance structures that might have taken shape, the predominant model, indemnity, was one that did not intervene in doctors'

relationships with patients or hospitals and did not interfere with their style of practice. Physicians suppressed . . . other financing models that would have placed doctors in a more subordinate position, models such as direct benefit insurance epitomized by the modern HMO. . . . The Great Society programs of the 1960s reflect the continuing institutional accommodation of physician interests. Cowed by fears of physician and hospital boycotts, lawmakers infused Medicare and Medicaid with a number of highly favorable structural elements patterned on the prevailing private insurance model. The dominating protectionist influence is codified in the programs' first words, which guarantee freedom from 'any supervision or control over the practice of medicine or the manner in which medical services are provided.' " Mark A. Hall, *Institutional Control of Physician Behavior: Legal Barriers to Health Care Cost Containment*, 137 U.Pa.L.Rev. 431, 446 (1988).

For these reasons, most attempts to solve the spending crisis are aimed at one or more of the traditional structural features. The next section describes these reforms in more detail, but it is useful to lay the groundwork here by describing the phenomenon of managed care. Much of this section

has been stated in the past tense since traditional insurance structures have given way to various components of managed care. These components are embodied to different degrees in the specific forms of insurance described below (HMOs, PPOs, DRGs, etc.), but it is possible to abstract from these particulars the three general features of managed care. Each responds to a key feature of traditional insurance. HMOs embody all three components to the fullest extent.

First, managed care limits the free choice of provider principle by contracting selectively with doctors and hospitals to form a network of providers, and then controlling movement within the network through "gatekeeping" physicians. HMOs require that subscribers enroll with a primary care physician in the network, who they must first consult in order to receive referrals to designated specialists. PPOs and point of service plans allow patients to go outside the network by paying higher deductibles and copayments.

The second feature is more intensive utilization review. No longer does insurance cover virtually anything a physician orders. Often, prior approval must be obtained from the insurer before initiating an expensive course of treatment. The third managed care feature is provider payment incentives. Rather than pay physicians and hospitals on a piece-work basis, HMOs and other insurers use payment incentives that reward providers for saving treatment costs. At present, virtually everyone with public or private insurance is subject to one or more

of these managed care components, even though less than half the insured population belongs to full-scale HMOs.

B. INSURANCE COVERAGE REFORM

To place recent reform efforts in perspective, we first review the broad spectrum of reform possibilities, ranging from the most to the least government involvement, by observing the systems of insurance present in other countries. Remarkably, this full range is also present already, in imperfect or compromised forms, in various portions of the complex and fragmented U.S. non-system.

1. SOCIALIZED MEDICINE AND THE BRITISH SYSTEM

The AMA often attacks reform ideas as socialized medicine, thinking that government take-over of the health care system lurks around virtually every corner. But seldom does anyone actually propose true socialized medicine. The purest example comes from the British National Health Service ("NHS"), where the government owns and operates the entire health care delivery system and guarantees comprehensive coverage to all of its citizens. This is generally considered to be a completely un-American approach, in view of the multitude of differences that exist between our two societies in political structures, cultural attitudes, and historical settings. Nevertheless, features of this approach are prominent in our current system, in the form of government hospitals. States have mental hospitals,

large cities have general municipal hospitals, the Veterans Administration and the military operate a large hospital and medical system, and the Indian Health Service maintains facilities mostly on reservations.

Listing these examples is usually sufficient to defeat this as an idea for comprehensive reform. Although some of these hospitals are among the finest in the country (e.g., Bethesda Naval), many are among the worst. Looking to England, where facilities are outmoded and overcrowded, confirms that government-run systems can be woefully underfunded. In our country, the condition and quality of government hospitals for the poor could be improved through increased funding, but as long as they serve primarily the poor, the necessary political constituency will always be lacking.

2. SINGLE PAYER, CANADIAN- STYLE INSURANCE

In contrast with socialized medicine is socialized insurance, in which the delivery of care remains in private hands but the government assumes the insurance or payment function. This is how the Canadian system is structured, and it is the essence of Medicare in our country. The purest reform idea is simply to extend Medicare to the entire country by essentially prohibiting private insurance, resulting in a single source of payment for all medical care. Although Canada, like England, has long waiting lines and underfunded facilities, this is not necessarily how the idea would take shape in the U.S.

if we were willing to support more generous funding. The potential compatibility of this idea with American values is demonstrated by the fact that Medicare is widely popular with the elderly.

Two other arguments speak in favor of single-payer reform. One looks to the savings in administrative costs. Competing private insurers generate a significant percentage of "wasted" spending on marketing, administration, and profits. Advocates claim these costs could be greatly reduced or eliminated through the streamlining and economies of scale of having a single, mandatory insurance plan. One measure of the potential savings is to look at the "medical loss ratios" for private insurers, that is, the portion of the premium dollar they collect that goes directly to paying for treatment. For private insurance, this ranges from 80 to 90 percent. Medicare, however, runs at a medical loss ratio in the mid-90s. The 10–15 percent savings that are available might be nearly enough by itself to pay for everyone without insurance.

The second argument is a moral one. A single-payer system promotes both vertical equity and horizontal equity. Vertical equity

speaks to who contributes. Horizontal equity speaks to who is eligible and what they receive. Private insurance requires at best that everyone contribute the same amount, but at worst requires those with the greatest need to contribute the most. On the receiving end, only those who pay are eligible for private insurance and its benefits are set by the level of contribution. In contrast, public insurance entirely disconnects

contribution from need and benefits. Contributions are based on the ability to pay, through the general taxation system. Benefits are determined entirely by need.

There are several compelling responses to these compelling arguments. The first is simply the pragmatic political observation that converting entirely to a single-payer system would mean that the entire health insurance industry would have to close shop. It would also mean moving hundreds of billions of dollars of expenditures from the private sector onto the tax-and-spend ledgers of the government. In an era of “no new taxes,” anti-entitlement programs, and eliminating budget deficits, this large scale conversion of 1/6 the economy is considered politically unfeasible.

The response to the administrative cost argument observes that not all these costs are wasted. Some portion goes toward holding down the costs of treatment, through managed care techniques. Another portion goes toward providing consumers more choice in the selection of insurance arrangements and benefits.

The moral argument raises the question of whether everyone should be entitled to the same level of insurance, determined entirely by need. It sounds agreeable to advocate equal access to health care, but we don’t currently maintain equal access to any

other fundamental social good such as housing, food, income, education or transportation. Instead, we attempt to provide a socially adequate minimum and allow people to make private purchases above the

minimum. A “two-tier” system of medicine violates some people’s egalitarian moral code, but many people are willing to accept some inequity if the bottom tier is at least decent. It is our insistence on state-of-the-art medicine that makes it difficult to afford any coverage for the portion of the population that remains uninsured. Relaxing the egalitarian standard could make decent coverage affordable for everyone. Insisting on strict egalitarianism would violate libertarian principles by requiring that upper limits be set on the care that the wealthy are free to purchase. Therefore, most people find some differentiation in access to be morally acceptable. This would mean allowing people to purchase private insurance that supplements or replaces public programs, which Canada does not allow. Medicare does allow this, but at present it discourages physicians from accepting more payment than Medicare allows, which frustrates the ability of Medicare patients to shop for the very best doctors.

3. EMPLOYER OR INDIVIDUAL MANDATES

A more realistic solution is to require either employers or individuals to pay for their own private insurance. An employer mandate was the heart of President Clinton’s failed proposal in 1993. To its credit, it attempted to build on the existing and popular employer-based private system, and it aimed to achieve economies of scale by covering most people in groups. Germany has

successfully achieved universal insurance through an employer mandate, as did Massachusetts and Hawaii prior to the ACA.

25

The primary objection is that increasing employers' costs will, similar to increasing the minimum wage, only hurt those we are trying to help by causing employers to eliminate a number of jobs at the margin. Another objection is that an employer mandate expands and perpetuates the distorting incentives created by exempting employer-provided insurance from personal income tax. This also neglects the self-employed.

Therefore, in 2007 Massachusetts became the first state to embrace the other alternative: an individual mandate, that is, requiring people to obtain their own insurance, either through employment or on their own, supported by government subsidies. At the same time, Massachusetts encouraged employers to continuing offering insurance, by imposing a "play or pay" tax that charges larger employers that opt not to provide coverage. This combination became the basis for the ACA's comprehensive health insurance reforms, although in 2019 Congress, at the Trump administration's urging, repealed the individual mandate aspect of the ACA.

The greatest difficulty with either an employer or individual mandate is structuring the subsidies that are necessary to make insurance purchase affordable for everyone. Employer requirements usually exempt smaller employers (those with 50 or fewer workers), only half whom provide insurance, which creates a disincentive to grow any larger. Subsidies for individuals are also difficult to structure. Family coverage costs roughly $20,000 a year, so even middle-income people need significant subsidies to

afford insurance, and lower-income need very substantial subsidies. But, the greater the subsidy, the more there is a disincentive to take on extra work or a better-paying job, since more earnings means fewer benefits. The solution in the Affordable Care Act is a sliding scale subsidy that extends all the way to four times the poverty level, which is roughly $100,000 for a family of four—well above the median family income of almost $60,000.

4. INCREMENTAL ALTERNATIVES

Republican lawmakers opposed the Affordable Care Act based both on its individual mandate and on its individual subsidies. Instead, they favored a set of more incremental reforms that attempt to shore up the places where insurance coverage is eroding the most rapidly, recognizing that this will not achieve universal coverage, but at least will move in the right direction, or will keep us from backsliding.

Examples of this incremental approach abound, both in public and private insurance. In the public system, examples include various expansions of Medicaid to cover children and pregnant women. In the private sector, there are various state and federal laws that make it easier for employers or workers to obtain and keep coverage. One of these laws, abbreviated COBRA, requires insurers to offer employees coverage at group rates for 18 months after they leave the job. Another law, abbreviated HIPAA, requires insurers after the 18 months expire to convert this group coverage to individual coverage,

but allows insurers to charge whatever they want. HIPAA, along with state laws, allow employees to switch insurers without undergoing a new pre-existing condition exclusion period, a concept known as "portability." This keeps workers with health problems from being locked into their current job.

This complicated and fragmented approach has the following limitations. First, prior to 2014, the purchase of insurance remained voluntary and unsubsidized, so many people continued to decline to purchase. Without a mandate to purchase, open enrollment attracts older, sicker subscribers, which makes insurance less affordable of younger, healthier ones, a phenomenon known as "adverse selection." Therefore, prior to the ACA, only a few states required insurers to offer individual insurance at community rates to all applicants, and those that did saw insurance premiums increase and the number of purchasers drop. More commonly, states created high risk pools for individuals who could not get health insurance, but this option usually cost 50 to 100 percent more than normal market rates, which already were very expensive. Thus, high risk pools were underfunded, and so tens millions of people went without insurance.

5. MANAGED COMPETITION

A final set of reform ideas known as "managed competition" cuts across this spectrum of proposals. It is adaptable to many of these approaches and therefore has been implemented in a number of settings. The essence of managed competition is a

voucher concept, in which subscribers can select from a range of subsidized insurance options in both the public and the private sectors, but in a monitored market environment that encourages

them to make a cost-conscious choice of the insurance that offers the best value for them. Unlike conventional Medicare and Medicaid, people would have a choice of insurance arrangements, some from the private sector. Unlike employment-based insurance, the choice would be made by the individual, not the employer. And, unlike either, the individual would bear the full marginal cost of the choice above the subsidy threshold.

Managed competition approaches vary quite a bit in their detail, but they look something like this. An employer or government program would give each person a voucher or subsidy that represents some or all of the value of a cheaper plan that offers decent coverage. The voucher amount would not vary according to the kind of insurance selected. If people pick a more expensive plan, they pay the full cost differential. They can select from any approved plan in the market. This maximizes choice and creates a strong incentive to shop for the best cost/quality trade-off. Managed competition does not necessarily equate with managed care, but it is assumed that, when faced with this choice, most people will opt for managed care.

This idea is widely adaptable. It is used by many large employers. Medicare was amended to create a Part C (called Medicare Advantage) which functions in roughly this fashion. Most prominently, the

Affordable Care Act creates an insurance purchasing exchange in each state for individuals and small employers. Under the ACA, people who receive private insurance subsidies from the federal government must use the new insurance exchanges to purchase their insurance, and smaller employers have the option of

purchasing through exchanges. These exchanges screen and monitor insurers to see that market rules are followed and that insurance benefits and prices are clear and comparable.

Whether this approach offers the optimal solution for structuring the insurance market remains to be seen. No idea is without potential problems. Difficult decisions are faced, for instance, determining which services should be included in the set of benefits that constitutes the mandatory minimum, for these benefits will then not be subject to market choice. Will the government be able to resist pressures to require coverage of expensive new medications or dramatic new innovations in diagnostic technology? Should it? Also, payments to insurers will have to be adjusted to reflect the relative risk status of different subscribers, in order to counteract insurers' incentives to avoid high risks. Otherwise, insurers could profit by engaging in selective marketing or poor service to high risk subscribers. Developing an accurate risk adjustment measure that cannot be gamed is a tall task that remains a work in progress. More fundamental is the concern that people will be confused and upset by all the choices and complicated information being forced on them, especially the elderly. In sum, the "managed" component of managed competition may involve much complexity

and government oversight if the system is to run smoothly and fairly.

C. INTERLUDE: ECONOMIC AND REGULATORY THEORY

The prior section reveals that two opposing political and economic theories—government regulation versus free markets—

influence different approaches for increasing access to health care, and we will see in the next section that the same is true for approaches to containing costs. This section is an interlude in which we discuss the theoretical framework for reform in more detail.

1. THE NEED FOR HEALTH CARE RATIONING

Our starting premise is that any successful reform effort must in some measure ration health care resources, that is, set limits on spending that deny some beneficial care, and in so doing allocate the limited resources among competing treatment needs. "Rationing" is a scare term used by those who oppose government involvement in health care. Therefore, law often prohibits government programs from openly or explicitly rationing care, pretending as if resources are limitless. But we know they are not, plus denying that they are simply leads to covert rationing. "Rationing" may seem an excessively harsh term for this inevitable reality, since allocation is inherent in society's use of *any* resource (given that all resources are limited to some degree), but explicitness is necessary here in order to shake us

from the romanticism that sometimes clouds the reality that even life-saving resources are not available in endless quantities. This is a disturbing reality because Americans have come to take medical miracles for granted and have come to expect the latest technological advances to be readily available to everyone. It is hard to conceive of Americans accepting, for instance, the practice that once prevailed in England where patients over a certain age seldom received kidney dialysis.

The symbolic values we attach to health care are illustrated by the contradictory attitudes we take toward saving "statistical" as opposed to "identifiable" lives. Daily, our society calmly suffers great statistical human tragedies such as highway deaths and the health effects of pollution or unhealthy lifestyles that could be reduced by increased tax funding for precautionary or educational measures, but when public attention is brought to an identifiable individual suffering from a present illness—say, a small child suffering from liver failure—we respond passionately with an outpouring of donations to support even extraordinarily expensive medical treatment. When the threat is imminent, there is a strongly-felt sense of injustice in a system that would allow a preventable death to occur, but not so for probabilistic risks of equal or greater magnitude. This paradox is one of the difficulties public policy makers face in engaging in more explicit health care rationing.

32

The need for health care rationing is better understood by viewing health care inflation as composed of two components: built-up waste and future waste. Current treatment patterns are infiltrated with instances of unnecessary care that have become ingrained in a fee-for-service practice style. We clearly need to trim the fat out of our present system, but medical science is continually producing technological advances that will quickly swamp the effect of such a one-time savings. In recent years we have seen the advent of magnetic resonance imaging and organ transplants. In the near future, we will witness the proliferation of artificial organ *implants*, even more exotic diagnostic machinery, and, not too far down the road, genetic therapies, all extraordinarily expensive yet promising medical advances. In order to accomplish meaningful reform, we

need some measure that will both tighten our current belts and resist future temptations.

Another useful way to discuss health care rationing is to identify the social and institutional levels at which it occurs. At the most global level, society can decide how much of its total resources to devote to health care versus other social needs, as the British Parliament does when it sets a budget for its National Health Service. At the most microscopic level, we can focus on which patients receive which treatments, as occurs when decisions are made about who has priority for scarce life-saving organ transplants. At the intermediate level, a rationing analysis can help decide which of the many branches of medicine are over or under-funded. This occurs, for

instance, in debates over whether health insurance should cover preventive medicine.

Each of the various cost-containment mechanisms we consider targets one of these levels and then spreads its effects throughout the other levels, resulting in rationing decisions of different forms and content. Before seeing how this happens in more concrete detail, we will explore other dimensions of rationing that are relevant at any of these levels.

2. THE ETHICS OF HEALTH CARE RATIONING

The inevitability of health care rationing requires us to confront two difficult questions: what are the criteria for rationing, and who should make the rationing decisions?

a. Rationing Criteria

The proper criteria for rationing medical care have been debated in a number of contexts. Most prominently, Chapter 6 discusses the criteria used to allocate a limited supply of life-saving organs for transplant. Later in this chapter, we discuss which areas of medical treatment should be covered by public and private insurance. In both discussions, there are common themes that point to contrasting ethical criteria.

One ethical approach is to use "neutral," non-medical criteria that avoid the tough value choices and trade offs. For instance, we could decide which of two candidates should receive an organ transplant by

tossing a coin or by picking the one who requested first, but these are usually viewed as arbitrary tie-breakers to be used only when more substantive criteria fail to provide a clear answer.

Which of several substantive criteria should we use? There are two basic approaches: medical need, and medical benefit. The two are not the same, since a patient who is in greatest need may be closest to death and so have the least chance of improving from the treatment in question. Imagine an organ transplant for a desperately ill patient who is likely to die in any event. More medical benefit can be gained by treating someone whose odds of surviving would increase from 50 to 100 percent than by treating someone whose odds would increase from 0 to 20 percent. For this reason, medical researchers have focused considerable efforts on developing objective measures of medical benefit.

One simplistic measure is age. Some ethicists, most notably philosopher Daniel Callahan, argue that health care allocation should favor the young over the old, under the view that this is

where it is likely to do the most good. But this rough proxy is not always accurate. Compare a severely deformed newborn infant with a healthy active 70-year-old who has a bacterial infection. A more nuanced measure of medical benefit is needed.

Medical benefit could be measured simply by the odds of saving a life, but some people live longer than others, so number of life-years saved is often used instead. But, some people might survive only to live a miserable, painful and disabled existence whereas

others might be restored to full health. To account for these possibilities, medical researchers often use a measure of benefit known as a QALY, for quality-adjusted life-year. This is a unit of measure that discounts the number of years added to life by a factor that reflects the degree of diminished quality of life. Thus, one treatment that might produce 10 years of life but with great pain and disability might receive a score of only 5 QALYs, and a different use of the same medical resource that produced 10 years of life in a permanent coma might receive a score of less than 1 QALY.

Making these comparisons is obviously a moral challenge, but if this can be done, then a measurement system would exist that would allow difficult comparisons to be made across not only different patients eligible for the same treatment, but also across different treatments and different diseases. For instance, a public insurance program faced with a difficult budget deficit could decide to allocate limited funds to prenatal care rather than to life-saving liver transplants if it documented that the prenatal care would produce more QALYs per dollar spent. An example of this approach from Oregon is discussed below.

Although QALYs do a better job than other rationing criteria, there are a number of compelling objections. First, there is no satisfactory way to make the quality adjustment that trades off life for various functional or mental impairments. One could ask people at random how many years of life they would be willing to lose in order to avoid certain conditions,

but can we really trust their answers, and who do we ask: only those who have experienced the condition, or only those who haven't? And, what do we do if answers vary widely? Even if these dilemmas could be solved, note that there is a disturbing utilitarian characteristic of QALYs, which is to equate 10 years of life for one person with 1 year of life for 10 people. Most everyone's intuition says the latter is a much greater benefit since more people are saved, but QALYs say the benefits are equal.

b. Rationing Decision Makers

It is certainly possible to fine-tune QALYs or other sophisticated techniques, but we can never expect these tools to resolve all the difficult dilemmas encountered in health care rationing. Ultimately, decisions must be made with imperfect information and unresolved social values and moral theories. How these issues are ultimately resolved therefore depends to a great degree on who the decision maker is. There are, broadly speaking, two alternative mechanisms for making rationing decisions: incentives and rules. Financial incentives, such as those inherent in various forms of prospective payment, can be directed to patients or providers to bring more cost-consciousness to their discretionary decisions. Alternatively, regulatory "command and control" mechanisms can be used to

dictate rationing decisions. Both approaches have considerable merit, but they each have serious drawbacks as well.

Directing financial incentives at patients means requiring them to have less comprehensive insurance

and to pay more out of pocket. This idea is discussed more below under the topic of medical savings accounts. It certainly has merit, but it faces these fundamental obstacles: Many patients simply cannot afford to pay out of pocket for a significant portion of their care. This is true not just for the poor, but also for those who are elderly or chronically ill, for whom predictable medical expenditures would consume far too much of their income.

For the generally healthy middle class, paying more out of pocket is feasible, but for many is not desirable. Nobel prize-winning economist Kenneth Arrow was the first to expound the "uncertainty" theorem that explains the prevalence of health insurance: Serious disease and accidents are expensive and dreaded events whose occurrence is not predictable; consequently, few people are able to plan rationally for the possible medical expenses simply by setting aside money or absorbing the costs when illness arises. People like comprehensive health insurance because they do not want to agonize over cost/benefit trade-offs when they are anxious about their health or that of a loved one. People generally are also not well informed to make cost/benefit trade-off decisions on their own, so they rely heavily on their physicians' advice. Therefore, it makes more sense to keep insurance comprehensive and direct financial incentives to physicians. They have the strongest influence on medical decisions as well as the best information.

These points are confirmed to some extent by a large-scale social experiment conducted by the

RAND Corporation in the late 1970s and early 1980s. These researchers randomly assigned people to different types of insurance. Those assigned to so-called "catastrophic insurance," which required them to pay much more out of pocket, indeed ended up spending considerably less on their health care. But these savings came primarily from decisions not to consult a doctor in the first instance. Once people consulted a doctor, they tended to incur the same costs regardless of how much they paid out of pocket. Also, analysis showed that the initial decisions not to seek care were not well informed. People did not accurately differentiate when they really needed to see a doctor and when then did not. Also, people who started the experiment both poor and sick ended up with measurably worse health status if they had to pay out of pocket, even though the research protocol gave them sufficient funds to pay for the uninsured part of their medical costs.

Financial incentives directed at physicians raise their own concerns, however. An imposing weight of ethical opinion supports the view that doctors should never allow their clinical judgment to be influenced in any manner by cost considerations. These ethicists insist that, because doctors act as fiduciary agents for their patients' welfare, they may not compromise optimal medical outcome in order to save money, especially if their patients are not responsible for any of the incremental costs of treatment decisions. *See, e.g.,* Norman Levinsky, The Doctor's Master, 311 New Eng.J.Med. 1573 (1984) ("physicians are required to do everything that they believe may

benefit each patient without regard to costs or other societal considerations").

This ethical perspective potentially finds expression in the law through the doctrine of informed consent. If physicians are called upon to make rationing decisions, the courts will be confronted with whether it is a violation of the duty of informed consent to withhold a treatment option that the doctor views as not worth its cost, without informing the patient that this resource constraint is being imposed. It might be objected that the duty of informed consent applies only to potentially harmful decisions to treat, but broader theories of informed consent would appear to cover treatment *refusals* as well. *See* Truman v. Thomas, 165 Cal.Rptr. 308, 611 P.2d 902 (1980). Whether these theories require disclosure of economic costs as well as of medical risks is an open question.

The ethical response is that, following the logic of the principal-agent theory just stated, it is at least ethically permissible, if not essential, for physicians to consider costs if doing so makes insurance more affordable or comprehensive. Patients might prefer their doctors to make the necessary cost/benefit trade-offs, considering that the alternatives are worse. This second-best or least-worst conclusion can be reached by employing hypothetical contract analysis such as that developed by philosopher John Rawls. It can also be reached through actual contracts, if patients choose their insurance fully aware of the role their physicians play in containing costs. This again underscores the merits of disclosing

the financial arrangements under which some physicians are paid.

Nevertheless, it is still troubling when physicians compromise optimal treatment not only to save money for the patient, but also to benefit themselves. This conflict of interest is in sharp contrast to the situation in England where salaried doctors make rationing decisions within a closed budgetary system. In that case, it can fairly be said that any money saved on one patient goes to help another whom the physician considers to be in greater need. In the United States, under HMO insurance and other forms of prospective payment, physicians or private insurers personally profit when they ration care. It is this conflict of interest that ethicists find most disturbing about provider-directed rationing. Because of it, many ethicists conclude that, if decisions to limit spending are imposed, they should come only from neutral sources external to the doctor-patient relationship, such as expert panels, citizen committees, or political representatives.

However, legislative or administrative regulation, which operates under the glare of the public eye, seems inherently incapable of making and enforcing the difficult decisions required in rationing health care. Regulators frequently capitulate to the intense pressure brought to bear by public and private interest groups. Even when regulators perform faithfully and evenhandedly, they necessarily do so according to a uniform standard, based on their view of the consensus opinion or prevailing values in the body politic. A monolithic response may be

inappropriate, however, in a pluralistic society where people have widely varying views of what is best or desirable in a particular case.

3. THE ECONOMICS OF HEALTH CARE RATIONING

a. Free Markets vs. Government Controls

Comparing incentives with rules raises not only ethical questions but also issues of economic and political theory. The focus of much of the debate over health care reform is on whether market-based solutions are preferable to government intervention. This theme arises in many different contexts, and few people believe that the debate can be resolved cleanly in favor of one side or the other in any absolute sense. Workable solutions will surely require combinations of both points of view.

Also, it is important to stress that the two opposing approaches are not mutually exclusive. As we discussed above with respect to financing reform, cost containment reforms lie on a spectrum of greater or lesser degrees of government controls versus free markets. For instance, we will see below that government programs can use market-like incentives rather than rules to "regulate" behavior. Similarly, private insurers subject to market forces can adopt internal rules that control physician and patient discretion. Also, markets can be allowed to function under rules and constraints monitored by the government. Rather than advocate one approach over the other, this section describes the themes and

characteristics of market versus regulatory approaches in the abstract, to help critique particular reform ideas as they are encountered later.

One such characteristic has already been noted, namely, that markets tend to permit a much broader range of value choices to be expressed than do government systems, which tend to impose a more uniform solution. On the other hand, decisions made in government systems are more visible and therefore more subject to

open debate. While this visibility lends to the legitimacy of the resulting decisions, it also makes it more difficult to make tough decisions, and it subjects these decisions to more political pressure by affected interest groups. Government also tends to implement its decisions through coercive, regulatory means, which people resist through evasion, circumvention or outright defiance. Enforcement therefore can be costly and ineffective. In contrast, markets operate through incentives and rewards, thereby achieving their objectives mainly through voluntary action.

These factors appear to point to market solutions as preferable, where they are available, which captures the political morality that prevails in other segments of the American economy. The difficulty is that market approaches may not be viable in health care because of its unique characteristics. The most obvious obstacle to the operation of competitive forces in the health care market is the existence of insurance. The fact that someone else is paying creates what is known as a "moral hazard" problem, meaning that the purchasing decisions of patients

and providers are distorted by the presence of insurance.

Second, many people have a fundamental philosophical opposition to pursuing a market-oriented mentality in health care. These critics maintain that such a mentality tends toward excessive commercialization and routinization of what should be a caring, individualized service. They view the hard-edged market as incompatible with the "soft," intangible values that we treasure in the healing arts, and they maintain that competition is incompatible with health care "as a caring rather than purely a curative activity, the goal of which is to reduce pain and anxiety and increase the

patient's sense of self-determination and quality of life." Rand Rosenblatt, *Health Care, Markets and Democratic Values*, 34 Vand.L.Rev. 1067 (1981).

Despite such criticisms, competitive forces are increasingly evident throughout the health care sector, as will be seen in the following section. To some extent, competition is focused directly on medical care decisions, with patients paying more out of pocket and with hospitals and drug companies engaging in direct advertising to consumers. For the most part, however, competition is focused at the level of insurance purchase. It is felt that, even if market forces do not work well for discrete items of care, these forces are capable of shaping the structure of insurance, which in turn sends institutional and monetary signals to physicians, hospitals and patients about how medical care decisions should be made. As noted above, even Medicare and Medicaid

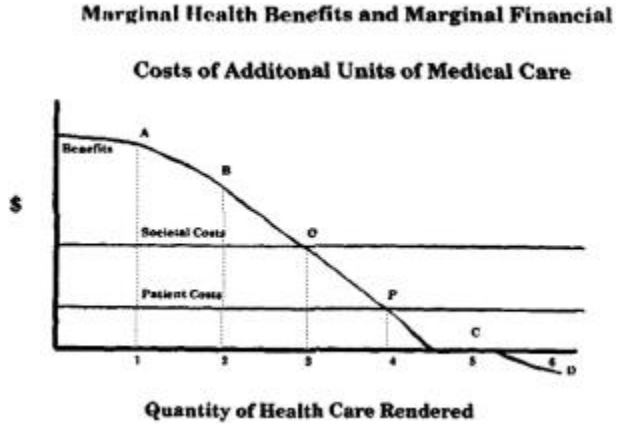
are turning to market-like voucher systems in which people receive coverage by enrolling with private HMOs.

b. Economic Theory

Even if an explicitly market-based approach is not adopted, still, economic analysis can help us understand what the ultimate objective should be for a successful approach to health care rationing. From an economic perspective, ideally, resources should be devoted to each medical procedure up to the point that their marginal costs equal their marginal benefits,[1] that is, the point where the last dollar spent produces no more or no less than a dollar's worth of additional benefit. A graphic representation may assist in understanding these concepts. (This analysis is adapted from Clark Havighurst and James Blumstein, *Coping With*

Quality/Cost Trade-Offs in Medical Care, 70 Nw.U.L.Rev. 6 (1975).) Figure 1 displays a hypothetical relationship between marginal increases in health care production and health care quality.

45



The vertical scale is the dollar value of health care (either its benefits or its costs). The horizontal scale represents abstract units of health care (days in hospital, doctor visits, drugs, x-rays, etc.). The "Benefits" curve shows the increment in societal health benefits that result from changes in the quantity of health care provided. Thus, at point A, when little health care is being provided, health care is very productive: there is a large return on each additional resource devoted to medical treatment. At some place down the line, though, we reach a point (B) of diminishing returns where additional health care expenditures rapidly become much less productive. As the line crosses the horizontal axis (point C), we reach the area of totally unproductive care, and below the axis is counterproductive treatment that produces a net medical harm (D).

The marginal benefits curve is in essence a societal demand curve for health care because it tells us, at

each level of production, how much society would be willing to spend on the last unit of service. If we then draw a hypothetical "Societal Costs" line to represent the marginal costs of the societal resources consumed by increasing units of health care, the optimal system would operate at the intersection of the two curves (O). (The cost curve is flat because of the simplifying assumption that each unit of health care has the same cost.)

We can also extend this economic analysis to the question of who should be the rationing decisionmaker. The diagram illustrates that where along the benefits curve our system tends to operate depends on the incentives that influence the behavior of the various actors that might make treatment decisions. For instance, under traditional insurance which guarantees providers full payment for any service they render, doctors will tend to function at C, because they seek to derive as much health benefit as possible, regardless of the cost. There is a risk that, if doctors become too focused on the financial rewards, or if they are too concerned about malpractice liability, they may tend to function at D, where providing these additional services leads to an actual decrease in quality of care. Patients may be somewhat more constrained by the payment of coinsurance and deductibles, but if these amount to only 20 percent of actual costs, their cost line will be 80 percent below the level of true societal costs.

We can predict a markedly different outcome under nontraditional reimbursement methods. Criticisms of HMOs and capitation payment suggest

they tend to operate closer to point B because decisions are made by corporate budget cutters who are insufficiently attuned to patients' needs. However, countervailing forces from market competition, malpractice exposure, regulatory oversight, or ethical standards might bring these incentives back in line. The difficulty is that we do not know whether these forces will adequately, excessively, or optimally counterbalance cost-cutting incentives since there is no ideal vantage point from which to measure what is ideal.

Therefore, this sort of analysis cannot generate a precise dollar amount for how much we should spend on any aspect of health care. Nevertheless, it establishes that an optimal health care system avoids spending money not only on unproductive or counterproductive care but also on beneficial care that is nonetheless more costly than it is worth—what might be called marginally unproductive care. One way to evaluate various proposals for containing costs, both regulatory and market-based, is whether they are capable of generating wise decisions about which beneficial treatments are not worth their costs.

D. COST CONTAINMENT REFORMS

1. REDUCING THE SCOPE OF INSURANCE

a. Practical Problems

The simplest way to contain costs is to reduce the scope of insurance. This can be done in three different ways: eliminate the bottom layer of coverage by

requiring patients to pay entirely out of pocket for most of their routine care. This is known as "catastrophic" or "high deductible"

insurance, meaning insurance doesn't kick in until medical expenses reach a very high level (say, $5,000–$10,000) in one year. The second approach is to eliminate the top layer of insurance by capping the total amount that will be paid in any year or over a lifetime. This is called "bare bones" insurance. The third approach doesn't have a common name, but it might be called "swiss cheese" insurance: coverage that is comprehensive for the services included, but which excludes entire categories or service, such as mental health or expensive organ transplants.

There are a number of objections to these approaches. First, they are not terribly popular with consumers. Even though economists tell us we have too much insurance, people have a persistent desire to be fully insured, for reasons explained above. Therefore, when these more economical options are made available, they typically sell less than their sponsors anticipated.

Second, some of these approaches may not save as much money as we might first imagine. Catastrophic insurance, for instance, appears to require that most medical care be paid for out-of-pocket since most people's medical bills never reach the catastrophic threshold in any given year. This is deceptive, however, since most medical spending is concentrated in the small number of people with very high cost illnesses. According to the "80/20" rule of thumb, the 80 percent of the healthiest people

account for only 20 percent of the total costs, whereas the 20 percent of the people with the highest costs account for 80 percent of total costs. For these sicker people, catastrophic insurance provides unlimited coverage once they exceed their deductibles.

Even below the deductible level, it is questionable how much out-of-pocket spending will be reduced. The theory of "consumer-driven" health care is that people will be forced to consider, for instance, whether it is worth it to have a sprained ankle x-rayed for the possibility of a fracture, to visit the doctor for a fever and sore throat, or to undergo an immediate operation for a mild hernia. Some people could not afford any of these options and so would suffer without needed care. To cushion the economic forces, federal tax law couples high-deductible insurance with "health savings accounts," which are tax-sheltered devices similar to IRAs from which people can pay their out-of-pocket medical expenses.

Another problem with catastrophic insurance is its tendency to aggravate "adverse selection." High-deductible insurance is much more attractive to the healthy than the sick. Therefore, its lower price reflects not only its greater efficiency, but also the underlying health status of the people who choose it. If high-deductible insurance were to become much more widespread, the cost of comprehensive insurance might skyrocket as it is left with all the sick people.

Other practical problems limit the potential for bare bones or swiss cheese insurance. Severely limiting the scope of insurance assumes we are

willing to deny care to those who choose not to, or are unable to, pay for it. But our society would likely resist the cruelty of denying someone life-saving care because they chose cheap insurance coverage. The "rescue ethic" that prevails in hospital emergency rooms and in intensive care units means that people with limited insurance can free ride on this humanitarian impulse.

b. Disability Discrimination and Other Legal Problems

Cost containment approaches that eliminate entire blocks of insurance coverage also encounter legal obstacles. Under public insurance, these decisions have been challenged as a violation of the statutory mandate to provide all "medically necessary" services or to administer benefits in a non-arbitrary fashion. *See* Beal v. Doe, 432 U.S. 438 (1977). In a number of cases this medical necessity mandate has proven to be an effective tool for obtaining Medicaid funding of controversial medical procedures. Some courts, for instance, have required Medicaid to cover medically necessary sex change operations. Rush v. Parham, 625 F.2d 1150 (5th Cir.1980). Most such litigation has focused on decisions by state Medicaid directors to restrict coverage for expensive organ transplants such as livers or hearts. In one decision, the court ruled that Medicaid could not cover liver transplants for some categories of patients but not others, since "to deny services arbitrarily and unreasonably to an otherwise eligible Medicaid recipient in this manner would be impermissible. There are some medical procedures, such as transplants, which Medicaid

participation does not obligate the state to provide. However, once these optional services are undertaken, they must be reasonably funded." Ellis v. Patterson, 859 F.2d 52 (8th Cir.1988).

Restrictions in the coverage of public insurance have also been challenged as disability discrimination under the American's with Disabilities Act (ADA), and its predecessor, section 504 of the Rehabilitation Act of 1973. 42 U.S.C. § 12101; 29 U.S.C. § 794. The latter applies only to publicly-funded programs and entities, but the former covers private businesses as well. An example of

how this law applies to health insurance comes from McGann v. H & H Music Company, 946 F.2d 401 (5th Cir. 1991), which arose prior to the effective date of the ADA. There, the court allowed a private employer to virtually eliminate health insurance for AIDS once it learned that one of its employees was infected with the HIV virus. Now, this is considered to be a clear case of discrimination based on handicap, since HIV infection constitutes a disability, as discussed further in Chapter 2. Bragdon v. Abbott, 524 U.S. 624 (1998).

Since virtually any health condition could be classified as a disability, does this mean that health insurance must cover everything in order to avoid disability discrimination? Not necessarily. In Alexander v. Choate, 469 U.S. 287 (1985), the Court ruled under § 504 that a state Medicaid program may cap covered hospitalization at a maximum of 14 days, even though it was clear that this would affect more severely patients who had more serious illnesses and

therefore who were disabled. The Court reasoned that the restriction was permissible because it did not specifically target a class of disabled individuals.

These rulings suggest that the cruder and less nuanced is the reduction in insurance, the safer it is under the disability laws. Eliminating entire blocks of service is permissible because this is neutral with respect to disability, but targeting specific illnesses can run afoul of disability discrimination. The difficulty this creates is that, in health policy circles, all efforts are directed at becoming more specific and fine-grained in decisions about what should and should not be covered by insurance. Reductions in coverage like

those in *Alexander v. Choate* constitute "rationing by meat ax," in the words of one commentator (David Eddy), because they eliminate services that might be lifesaving for some patients but allow payment for other conditions where the service might be completely unnecessary. Health policy researchers much prefer to make these cutbacks based on the relative cost effectiveness of the medical service as applied to various medical conditions.

The emerging science of cost-effectiveness evaluation has generated a great deal of controversy. Many ethicists adamantly oppose placing a monetary value on life and health, but cost-effectiveness analysis avoids this by attempting to value alternative methods for incremental medical gains. Without having to decide how much a life or a year of life is worth, it is still possible to evaluate which of several alternative treatments produces the greatest

gain, and which does so with the least cost. It is also possible to measure, at the margin, how much incremental gains in life-expectancy cost as advances in medical technology occur. Thus, researchers once calculated that performing a Pap smear every three years to detect cervical cancer cost $12,000 per additional year of life expectancy, but increasing the frequency of testing to once a year cost $930,000 per additional life-year saved. This does not tell us which frequency of testing is correct, but it does help us decide where cuts in funding can be made by doing the least amount of harm, when cuts are necessary.

The most prominent example of using cost-effectiveness techniques for medical technology assessment is the rationing scheme adopted by the state of Oregon to allocate its limited

Medicaid funding in a way that allows it to cover all needy people. Prior to 2014, most states covered all medical services but were able to afford only about half of people below poverty. Oregon chose instead to cut back on which services it covered so that it could afford to cover everyone in poverty at whatever level funds allowed. In order to decide how best to cut back on covered services, Oregon created a task force to evaluate the relative effectiveness of the full range of medical services, classified into about 700 categories of conditions and treatments. By eliminating those services that produced the least medical benefit, Oregon was able to afford coverage for only about 590 of these items. This approach was challenged by federal Medicaid administrators as a violation of the ADA, however, because some of the excluded items targeted conditions that constitute classic

disabilities, such as liver transplants for alcoholics, and intensive care for severely deformed or premature newborns. Oregon officials had to make extensive revisions in order to salvage their innovative plan.

This solution may appear to work an acceptable compromise between the need to ration scarce medical resources and the concerns over disability discrimination, but in fact it does not. Disability discrimination concerns are still at the core of any effective rationing mechanism even if the most visible categories are avoided. This can be seen best by recalling the discussion above, in section C.2.a, of using “quality-adjusted live years” (QALYs) to gauge the effectiveness of different medical treatments. This measure responds to the concern that mere life expectancy is too crude a measure, since some lives that are saved or lengthened may leave people in severe pain or in a permanent

coma. But, adjusting the measure of value according to the quality of life that results from treatment would provide fewer resources to those who end up with greater disabilities. Once again, the more refined and sophisticated are the tools for medical technology assessment, the more likely they are to run afoul of disability discrimination laws. It may be possible to reconcile some uses of QALYs and other cost-effectiveness measures with these laws, but the tension between these two social objectives still lies at the core of rational health planning. The same is true for more covert ways of rationalizing restrictions in treatment, such as the "futility" concept discussed in Chapter 7.E.

2. RIGOROUSLY REVIEWING THE NECESSITY OF CARE

An alternative to reducing the scope of insurance is to keep insurance comprehensive but to impose a more demanding standard for when services are necessary in particular cases. Most public and private insurance restricts coverage to services that are "medically necessary" and not "experimental." Over the past decade, insurers have become much more demanding about when services are established and appropriate for specific patients or conditions. This has resulted in a spate of litigation concerning the proper interpretation of these coverage terms.

In one line of cases, courts have been very reluctant to allow insurance companies to deny reimbursement based on their determination that the services are not medically necessary. One leading decision is Van Vactor v. Blue Cross Assoc., 365 N.E.2d 638 (Ill.App.1977), where the court reasoned that, because "the term 'medical necessity' is subject to various interpretations, . . . there was sufficient evidence to warrant [the trial court's

conclusion] that the insured was justified in relying on the good faith judgment of his treating physician" in ordering hospitalization for oral surgery. The court was much more adamant in Mount Sinai Hosp. v. Zorek, 50 Misc.2d 1037, 271 N.Y.S.2d 1012 (1966), which reversed a Blue Cross refusal to pay for three weeks of hospitalization to administer a severe weight reduction diet. The court held that "only the treating physician can determine what the appropriate treatment should be for any given condition. Any

other standard would be intolerable second-guessing, with every case calling for a crotchety Doctor Gillespie to peer over the shoulders of a supposedly unseasoned Doctor Kildare."

These holdings have not found uniform acceptance. Other courts have been willing to enforce reimbursement denials, particularly where the insurance policy explicitly states that the insurer has final authority to determine medical necessity. Sarchett v. Blue Shield of Cal., 233 Cal.Rptr. 76, 83, 729 P.2d 267 (1987) ("it is unlikely that any insurer could permit the subscriber free selection of a physician if it were required to accept without question the physician's view of reasonable treatment and good medical practice"); Lockshin v. Blue Cross of Northeast, Ohio, 434 N.E.2d 754, 756 (Ohio App.1980) ("a function, basic to the insurer, is the right . . . 'to determine whether . . . [a] claim should be allowed or rejected' ").

The outcome of these cases is affected by a number of factors. First, the standard of review is greatly affected by which court the case is in. State courts deciding garden variety contract interpretation cases are much more prone to give the benefit of the doubt to the patient. However, most of these cases are now decided

in federal court under a standard of review more lenient to the insurer. This is as a result of ERISA preemption. ERISA is the federal statute discussed in Chapter 5.C which preempts certain bodies of state law with regard to insurance that is provided as an employment benefit. Here, the effect of ERISA is to force insurance coverage disputes into

federal court, to be decided under principles of federal common law. Following Supreme Court precedent, courts defer to insurers' judgment when the contract declares that insurers have the authority to interpret its terms. Firestone Tire & Rubber Co. v. Bruch, 489 U.S. 101 (1989).

The other change that has affected this area of litigation is the shift from retrospective claims review to prospective utilization review. In the past, coverage disputes arose after treatment was rendered, when claim for payment was submitted to the insurer. Now, insurers require that physicians and patients obtain their advance permission before undertaking an expensive course of treatment. This was intended to avoid the unfairness of refusing to pay after a patient has relied on a physician's advice, but it has created a different hardship. Before, denial of coverage only affected payment, not treatment, but now the effect is to tell the physician not to treat at all. This has brought to the courts very high stakes, life-and-death decisions in which they must decide quickly whether to issue an injunction to order an experimental procedure that might save the patient but which has not yet been clearly proven effective.

The undesirability of resolving medical appropriateness issues in this fashion has prompted a search for superior dispute resolution

processes, ones that are quicker and easier to access, that avoid the bias of delegating all authority to insurers, and that bring more expertise to bear than exists in the courts. These issues are likely to be addressed in efforts pending at the time of this writing to enact

patient protection legislation for managed care insurance.

This area of litigation has also caused insurers to examine the processes by which they conduct utilization review. Typically, they hire nurses, and sometimes lesser trained staff, to take calls over a toll-free number. Treating physicians or their assistants describe the patients' condition and the proposed treatment, and reviewers compare these indicators with screening criteria the insurer maintains in computerized practice guidelines to determine medical appropriateness. If the computer flags a treatment request as questionable, then the request is reviewed by a physician, often a general practitioner with no special expertise in the medical field in question.

Physicians complain this is a poorly-designed process that nit-picks their professional judgment and dictates patient medical decisions without proper credentials or investigation. They insist that utilization review criteria should be publicly disclosed, and that the personnel should have better credentials. So far, however, regulators and courts have been generally accepting of this process. Regulatory authorities set minimum standards, but so far they have required only that personnel be licensed and that the process provide a quick response. Similarly, in litigation, courts have observed that insurers only purport to make payment decisions, not treatment decisions. Physicians are free to, and perhaps are

required to, render care they think is necessary even if they won't be paid.

59

Physicians can also assist patients in appealing coverage decisions they think are wrong. *See* Wickline v. State, 228 Cal.Rptr. 661 (Cal.App.1986).

Even though detailed utilization review is now a permanent feature of the health care landscape, no one believes that it alone can solve all the tough problems. It is an expensive and clumsy process that reaches only major decisions whose cost warrants special scrutiny. Also, it is limited by the capacity of computerized protocols to capture the complexity and nuance of medical decision making. There are more than 10,000 medical diagnoses and 10,000 different treatments. Devising thorough and accurate guidelines for each of the millions of possible combinations based on solid empirical evidence is an impossible undertaking, an insight that is captured in the slogan that medicine is as much art as it is science. Therefore, a broad range of medical practice will necessarily remain subject to individualized professional discretion. To reach this portion of decision making, insurers have adopted a variety of payment innovations that seek to make physicians and hospitals more cost conscious.

3. REFORMING PROVIDER PAYMENT

The multitude of payment methods for doctors and hospitals can be better understood if they are arrayed in a spectrum from the most open-ended to the most encompassing. This spectrum looks at the structure of the payment method rather than the absolute amount of payment. At the least restrictive end is fee-for-service payment which reimburses providers for

each item of care they deliver. At the most restrictive end is salaried employment for physicians and a fixed, global budget for hospitals, according to which each receives only a single, fixed amount that does not vary in any respect according to the number of services provided. Stretching between these two extremes are various intermediate versions that have greater or lesser aspects of variability or "prospectivity."

In this context, prospective payment means a payment method that fixes the amount of reimbursement in advance of treatment and therefore imposes on the provider some degree of risk or potential for profit. If actual treatment costs are lower than the fixed amount, providers profit, but if more, they absorb the loss. Prospective payment thus attempts to replicate market-like forces by giving providers incentives to economize. Prospective payment can take the form of a fee schedule, in which fee-for-service reimbursement is preserved but the amount per service is fixed by the payor rather than set by the provider. Or, prospective payment can take the more global forms of an annual salary or annual hospital budget. In this section, we will explore the more prominent forms of prospective payment developed by public and private insurers.

a. Medicare Prospective Payment

In 1983, the federal government entirely reformed the traditional Medicare system of retrospective, cost-based reimbursement for hospitals, replacing it with a new "prospective payment system." This

provides a good starting point for understanding the complexities of designing an efficient and fair set of payment incentives. First, consider the possible units of service for which hospital payments might be fixed in advance. If Medicare paid hospitals a fixed amount for each day that a patient is hospitalized, hospitals would continue to have an incentive to treat patients as long as possible. If Medicare instead paid hospitals a fixed amount for each Medicare patient admitted, it would undercompensate hospitals that, because of their location or specialization, treat relatively sicker patients and overcompensate community hospitals in suburban locations that cater to relatively healthy patients with comparatively minor ailments. Therefore, the solution Congress chose was to fix an amount for each patient admitted according to the patient's diagnosis.

This diagnosis-based method of reimbursement is known as the "DRG" method of payment—for "diagnosis-related groups." Specifically, Medicare takes all medical diagnoses and groups them according to their relative medical resource consumption. Each of these approximately 480 groups is assigned a weighting factor, which is then used to adjust the average cost of treating all Medicare patients. For instance, the DRG for "all major chest surgeries" carries a weight of about 3.0, reflecting an expensive hospitalization. "Other respiratory surgeries" are placed in one of two groups according to whether they are accompanied by complicating conditions, and these groups are assigned lower weights of about 2.5 and 1.5. Assuming an average cost per case of $10,000, the

first DRG mentioned would pay the hospital $30,000, the second $25,000, and the third $15,000. In each case, this is all the hospital

gets, even if the patient has multiple other conditions, but the hospital keeps this entire amount, even if the patient can be treated and discharged much easier and quicker than normal.

The DRG form of prospective payment is not unique to Medicare. It has been widely adopted by state Medicaid programs and private Blue Cross insurance. Despite this success, few if any health policy analysts are convinced that DRGs will produce lasting reform. First, the DRG system has only limited reach. It covers only *in-patient hospital services.* Excluded are physicians' fees and outpatient facilities. The seriousness of these limitations is documented by the one-third increase in Medicare outpatient expenditures that occurred during the first two years of DRGs. As a result of this shift in services to more unconstrained reimbursement environments, DRG-type payment methods have been developed for a number of additional facilities, including nursing homes, home health agencies, and hospital-based outpatient services. It is questionable, though, whether DRGs are capable of being extended to the core of medical practice:

> paying physicians is far more complicated [than paying hospitals]. When developing a hospital-payment system for Medicare, one must handle 11 million admissions to 7000 hospitals for 475 diagnosis-related groups. Those numbers pale in comparison to Medicare's 350 million claims

> from 500,000 physicians for 7000 different procedure codes. Moreover, whereas hospitals can average their gains and losses under a prospective payment system across many cases, physicians' smaller caseloads and greater specialization make such averaging much more risky for them. These differences

> mean that improving the way Medicare pays physicians will be vastly more difficult, both analytically and administratively.

Roper, *Perspectives on Physician-Payment Reform,* 319 New Eng.J.Med. 865 (1988).

Owing to these considerable difficulties, the government appears to have abandoned any attempt to design physician DRGs. Instead, it developed a modified system of fee-for-service reimbursement known as a "relative value scale." The version Medicare developed is known as a "resource-based" relative value scale (RB-RVS for short) because it attempts to achieve some degree of parity in the amount that physicians charge for various services by measuring the relative costs of each service according to the time, mental effort, and technical skill required, as well as differences in the costs of malpractice premiums and specialty training. Such a system does little, however, to alter the existing incentives that continually drive up the volume of physician services.

The ample opportunity that DRGs present for manipulation creates grounds for questioning their effectiveness even within their present ambit. DRGs pay hospitals for each patient admitted; therefore,

they carry the potential to induce unnecessary admissions. Even for patients who do need hospitalization, DRGs create an incentive to exaggerate the diagnosis. Hospitals encourage doctors to place their patients in higher-weighted DRGs, a phenomenon known as "DRG creep." Moreover, hospitals attempt to "unbundle" medical treatment so that services are placed in the maximum number of reimbursable categories. For instance, hospitals can receive extra reimbursement by performing diagnostic workups prior to

admission and by transferring patients to hospital-owned nursing homes after discharge. This sort of activity is wide-spread, which partly explains why outpatient costs mushroomed at the same time DRGs were introduced.

These abusive practices aside, DRGs may prove ineffective even at their core because of a variety of counteracting incentives and barriers inherent in our health care system. Hospitals may have difficulty responding even in a socially beneficial way to the economizing incentives of prospective payment because hospital costs are almost entirely dependent on the treatment decisions of physicians, who continue to be paid on an inflationary, fee-for-service basis. Legal doctrine detailed elsewhere in this book amply protects doctors from outside, financial influence on their clinical judgment. *See* Chapters 5.A (unlicensed and corporate practice of medicine), 5.E (referral fee prohibitions), and 3.B (access to the medical staff). Even if such influence could be brought to bear, malpractice laws and the forces of

quality competition might deter any lowering of the intensity of service.

Other, more subtle, potential problems inhere in the DRG system's attempt to pay the average cost of treatment for each patient. Although this payment is adjusted according to the patient's diagnosis, each patient is assigned to only a single DRG and receives only a single payment[2] based on the principal diagnosis, regardless of the severity or number of illnesses the patient has. The very broad range of illness severity within each diagnostic category creates a strong incentive for hospitals to admit less seriously ill patients. This may lead to overtly discriminatory

admissions practices, the validity of which can be tested under the principles developed in Chapter 2. More troubling, though, hospitals may find subtle methods for "case-mix management" that pass any possible legal scrutiny—techniques such as eliminating burn units, emergency rooms, or other services that tend to attract severe cases, or by relocating from low-income, inner city population centers where patients as a group tend to be sicker. Over time, these incentives might further exacerbate the serious problems of access to health care that certain population groups already face.

The DRG system responds to these problems to some degree by varying the basic payment rate

according to various additional factors that reflect the type and location of the hospital. For instance, each of these hospital types receive additional increments to compensate for their higher costs or greater social mission: teaching hospitals, hospitals with a disproportionate share of low-income patients, hospitals that face higher wage costs, and rural hospitals that are the only facility in the area. But, each of these refinements adds greater complexity and administrative controls to what initially was thought to be a simple incentive-driven system.

As a result of these many criticisms, it is clear that DRGs do not hold the final solution to the health care cost crisis. Congress is now attempting to implement less administratively-intensive strategies that rely on a system of competing private HMOs and health plans, and that encourage providers to structure themselves into "accountable care organizations" that receive bundled payments covering both hospitals and doctors. Whether in fact

these systems are simpler, fairer, and more efficient remains to be seen.

b. Capitation Payment and HMOs

A different form of prospective payment known as "capitation" has taken hold in the market for private insurance. The most common embodiment of capitation payment is the HMO, which stands for health maintenance organization. Health policy analyst Paul Ellwood is credited with coining this term in 1970 to describe what had previously been known more descriptively as "prepaid group

practice." The new term is meant to emphasize the HMO's focus on preventive care. This focus springs from the fact that providers are paid a single amount per patient enrolled ("per head," hence, "capitated") to cover all of the medical needs for a prescribed time, usually a year. The term HMO is meant to emphasize the positive incentive this creates to keep patients healthy rather than the incentives under traditional fee-for-service payment which pay doctors more the sicker their patients are.

Capitation is potentially a powerful force for cost containment because it dramatically reverses the traditional financial incentives created by health insurance. HMOs profit by treating less rather than more. HMOs also bring cost consciousness to bear precisely at the point of the most informed treatment decisionmaking: the attending physician. In essence HMOs combine the treatment function and the insurance function into a single entity, in contrast with "third-party" reimbursement.

Precisely how capitation incentives affect medical judgment is not yet well understood. In part, this is because the precise financial arrangements within HMOs are complex and vary widely. Capitation defines the method of payment to the HMO as an entity, but it might choose to pay its physicians and hospitals in any of a number of other ways, including salary, discounted fee-for-service, and fee-for-service with various penalty or reward systems that encourage economizing. (These are called "withhold pools" because they usually operate by withholding a portion of the contracted payment to either pay out

or forfeit at year end based on whether performance goals are met.)

These methods of physician payment tend to correlate with different forms of HMOs. In group or staff model HMOs, doctors practice together in the same setting, whereas individual practice associations ("IPA") and network model HMOs are a looser contractual association of a larger number of doctors who maintain practices in their individual offices. Staff HMOs usually employ their physicians on a salaried basis. IPAs typically compensate their physicians on a discounted fee-for-service basis supplemented with the bonus/penalty arrangements just described. Group and network models employ a variety of physician payment techniques, the most notable of which is capitation. When HMOs make capitation payments to their doctors and hospitals, they shift much of their financial risk directly to the providers that recommend and render the care.

In a typical arrangement, the HMO might keep 20 percent of the capitation payment it receives (to cover sales expenses, administrative overhead, and profit), and then split the remainder

between contracting hospitals and physicians. For instance, each primary care physician might receive 40 percent of the capitation payment for each patient for which she is responsible, to cover all physician services and pharmacy costs, including the costs of specialists to whom they refer their patients for more complex problems. The remaining 40 percent might then be set aside in a pool to pay for hospitalization costs authorized by this physician. If there are shortfalls

in the hospitalization pool, these might be absorbed by the HMO or partially deducted from the primary care physicians' pay. Alternatively, under an approach known as "global capitation," a large multi-specialty physician group might accept the full 80 percent capitation payment and then further contract "downstream" with hospitals and specialists.

Whatever the precise arrangement, the consequence is that physicians have strong incentives to economize both by promoting health and by minimizing treatment for the sick. HMOs in fact spend much less on hospitalization than do conventional providers. They use inpatient facilities as much as 40 percent less than traditional fee-for-service practitioners without a large accompanying increase in the ambulatory treatment. And, as discussed in Chapter 5.B.3, they do so without demonstrable harm to patients and, by some measures, better results. This financial and clinical success has spurred rapid growth in the industry. HMOs now account for over half of all privately insured people and they are becoming much more prominent in Medicaid and Medicare. Many state Medicaid programs have turned, in whole or in part, to HMOs for health care delivery to the poor. And Medicare promotes HMO enrollment as an alternative.

This fundamental shift in the structure of provider reimbursement raises a host of crucial legal and social issues, most of which are explored elsewhere in this book. Here, we focus on the limits to which

HMOs can attempt to influence physicians with financial incentives. Capitation creates a serious conflict of interest between the patient's best medical interests and the physician's economic interest. Many ethicists adamantly oppose any form of financial inducement to bedside rationing, that is, any incentive that would cause physicians to compromise optimal care on account of costs. They reason that this would fundamentally compromise physicians' ethical role as devoted patient advocates, would undermine the trust that is essential to successful therapeutic encounters, and would lead to abuse. Other ethicists and commentators respond that, since rationing is inevitable, some rationing decisions are better made through the nuanced, discretionary, and patient-sensitive judgment that is possible only at the bedside. They reason that, if bedside rationing is to occur, some forms of financial motivation are permissible, if not ideal, and the legal response should be regulatory, not prohibitory. *See* Mark A. Hall, *Rationing Health Care at the Bedside*, 69 N.Y.U. L. Rev. 693 (1994).

This basic conflict in perspectives underlies many branches of emerging health care law. It influences malpractice law, informed consent obligations, and patients' rights legislation as they affect HMOs. Here, we will focus on regulatory law that directly limits these financial incentives or that requires their disclosure. The primary source of law is a statute and set of implementing regulations directed to Medicare or Medicaid patients in hospitals

and HMOs. 42 C.F.R. § 417.479. For hospitals, this law takes a prohibitory approach. It flatly bans any "payment,

directly or indirectly, to a physician as an inducement to reduce or limit services provided with respect to [Medicare or Medicaid patients who] are under the direct care of the physician."

For HMOs, however, this law takes a more permissive, regulatory approach. It bans only financial incentives directed to a single patient and that are designed to limit care that is medically necessary. Other incentives are allowed, however, according to various parameters that affect their strength and immediacy. For instance, financial incentives that affect only physicians' time and effort and not their income, such as capitation for primary care physician services, are not restricted. Also exempt from regulation are incentives that are pooled across a group of doctors who treat a large number of patients (25,000) so that withholding care for any one of them does not impose any substantial penalty on the responsible physician, yet the group has an incentive to economize on all care. Finally, the Medicare regulations require certain safeguards, oversight measures, and "stop-loss" protections for incentive arrangements that put more than 25 percent of a physician's compensation at risk.

This complex regulatory approach, which permits some incentives, bans some, and limits some, is likely to become the legal norm as these rules are adopted by states and applied to private insurers. In addition to whether these incentives are permissible, the law must determine the extent of their disclosure. As discussed in Chapter 2.B.3, plaintiffs' lawyers are beginning to argue in malpractice litigation that

failure to disclose financial arrangements with HMO physicians that create a conflict of interest violates the fiduciary nature of the treatment relationship and constitutes lack of informed consent to treatment. A number of states require by statute that these and other disclosures be made to HMO subscribers.

4. PUBLIC UTILITY REGULATION

DRG prospective payment and HMO capitation create passive, market-based reimbursement incentives to control medical costs. In contrast, a much more heavy handed, "command and control" regulatory approach could be employed. This is the approach embodied in the certificate of need laws (CON) enacted in the 1970s (discussed in Chapter 3.A.2), which require agency approval before making capital expenditures or offering new services. This approach is also embodied in so-called "all-payor" hospital rate regulation, which a number of states (mostly in the Northeast) instituted in the 1970s and 1980s. All-payor rate regulation takes a prospective payment method such as DRGs or global budgeting and applies it uniformly to all sources of payment, including private insurance and out-of-pocket payment. In essence, hospitals are treated like regulated monopolies, similar to electric companies, local telephone service, and other public utilities.

This approach is currently not in vogue. All-payor rate regulation has proven ineffective, and it has been repealed in almost every state that tried it. The federal requirement for state CON laws has also been

repealed. Nevertheless, elements of the public utility approach remain. CON laws still exist in most states, and prospective payment methodologies imposed by Medicare and Medicaid maintain a considerable degree of federal and state rate regulation. Whether the public utility approach will be revived in the future remains to be seen. This is how hospitals are regarded in Canada, England, and other countries with universal health care systems, and it is how many critics allege that the Affordable Care Act treats health insurers.

E. RECAPITULATION

Several themes emerge from the policy discussion in this chapter that have central relevance to the doctrinal topics addressed in the following chapters. First, studying the legal infrastructure that buttresses the traditional institutions and relationships in medicine will help us better understand the causes of the crisis in health care spending. Second, in a time of tremendous ferment, experimentation, and change, it is critical to recognize the numerous challenges this upheaval will present to conventional legal thinking as it struggles to adapt past doctrine to the new circumstances. Third, there is a mind-boggling array of legislative and market strategies for containing costs and expanding access, and an equally daunting array of anticipated responses from insurers, providers, and patients. While it is helpful in understanding this complex tapestry to sort reform techniques into market-based versus regulatory approaches, this does not resolve which approach is best or is most

likely to take hold. Virtually every reform technique that has been conceived exists in some form or fashion in the highly fragmented

and hugely complex set of institutions, laws, and policies that make up the American health care "system."

[1] "Marginal," in this context, means the incremental cost or benefit attributed to each additional unit of service, as opposed to the *aggregate* value for all services combined, or the *average* value per service. For instance, a hospital's marginal cost for treating its first patient is very high—the cost of the entire hospital—but the marginal cost for its second patient is quite low, since the hospital is already built, the nurses employed, etc. Similarly, the marginal benefit to a patient in cardiac arrest of the first few minutes of life saving care is quite high, whereas the marginal benefit of the last few minutes of a two-week hospital stay is negligible.

[2] An important exception to this statement exists for cases classified as "outliers," patients whose costs of treatment lie far outside the normal range. Hospitals receive an additional payment for outliers; however, this payment covers only a portion of the costs of extended care.

CHAPTER 2
THE TREATMENT RELATIONSHIP

Any treatment of health care law faces the daunting task of deciding which topics to present and in what order. Health care law, as it has developed over the past four decades, has become an unwieldy collection of disparate areas of doctrine and public policy. Accordingly, some lawyers and scholars maintain there is no unifying structure or core set of ideas that qualifies this as a coherent and integrated body of legal thought and professional practice, other than the happenstance that each topic involves doctors, hospitals, or health insurance in some way. We agree this field has not yet jelled in the way that classic first-year subjects have, but we nevertheless see interlaced throughout these disparate topics several organizing principles or themes that potentially explain not only what makes these disparate parts cohere, but also why that coherence distinguishes health care law from other bodies of law.

One of these of organizing themes, and perhaps the most prominent, is the set of attributes that make the medical enterprise uniquely important or difficult in the legal domain. Health care law is about the delivery of an extremely important, very expensive, and highly specialized professional service. If anything distinguishes health care law, it must be the unique aspects of the treatment encounter viewed from both sides of the doctor-patient relationship. Health care law in each of its branches must take account of the phenomenology of what it is

to be ill, to seek treatment, and to be a healer. These human realities are permanent features that distinguish this field from all other commercial and social arenas and alter how generic legal doctrine and conventional economic and political theories respond to its issues and problems.

Accordingly, this chapter focuses on the doctor-patient relationship. The first part explores the legal rules that govern the *structure* of the treatment relationship: the duty to treat, and the formation and termination of the relationship. The second part looks more into the *content* of the treatment relationship. It surveys the legal doctrines and policies that flow from the fiduciary nature of the relationship, including confidentiality, informed consent, and the contractual modification of treatment obligations.

A. DUTY TO ACCEPT AND TREAT PATIENTS

A patient's right to receive treatment differs between paying and indigent patients. Paying patients, of course, have greater access. But certain obligations to treat are independent of the ability to pay, and for paying patients there remain some barriers based on race, disability and other factors. Thus, the ability to pay is not the primary issue in the following discussion. Nevertheless, access to medical care by the indigent is the subject of both explicit and implied attention in much of what follows.

1. DOCTORS

a. The "No-Duty" Rule

A doctor is generally under no duty to accept patients, regardless of the seriousness of their condition, their ability to pay, or the physician's basis for refusing. In the seminal decision, which is still

regarded as stating "good" (*i.e.,* prevailing) law, the court affirmed the dismissal of a suit for damages on behalf of a would-be patient who died when a physician refused to treat her—notwithstanding that the doctor had been her family physician in the past, was available to render care (and aware that other physicians were not), was told she was now seriously ill and relying on an expectation of treatment, gave no reason for the refusal, and was offered payment. Hurley v. Eddingfield, 156 Ind. 416, 59 N.E. 1058 (1901). The court reasoned that a physician's traditional freedom to select patients remained unaltered by the advent of state licensure law, which imposed no obligation on a physician "to practice at all or on other terms than he may choose to accept." A half century ago, a Texas doctor refused to attend to a pregnant "Negro girl in the emergency room having a 'bloody show' and some 'labor pains' "; as a result, the baby lived only 12 hours. The court had no qualms about pronouncing:

> Since it is unquestionably the law that the relationship of physician and patient is dependent upon contract, either express or implied, a physician is not to be held liable for arbitrarily refusing to respond to a call of a

> person even urgently in need of medical or surgical assistance provided that the relation of physician and patient does not exist at the time the call is made or at the time the person presents himself for treatment.

Childs v. Weis, 440 S.W.2d 104 (Tex.Civ.App.1969).

This "no-duty" rule tracks the historical absence, in American tort law, of any legal obligation to aid strangers in distress. Professional medical ethics reflect a similar policy: "Even the Hippocratic Oath,

by which every doctor is morally bound, assumes a pre-existing relationship of patient and physician, which relationship in its inception is basically contractual and wholly voluntary," Agnew v. Parks, 343 P.2d 118, 123 (Cal.App.1959), and the AMA Principles of Medical Ethics leave a physician free "to choose whom to serve" (though they recognize an exception for emergencies).

Implicit in these articulations of the "no duty" rule is the axiom that, where a physician/patient (or hospital/patient) relationship *does* exist, there *is* a legal obligation to treat. This duty to treat is fiduciary in nature (sec. B.1.) and persists until the relationship is properly terminated (sec. A.6). Since the formation of the treatment relationship is foundational to the entire range of issues that make up law and medicine (including malpractice and most of bioethics), what constitutes "formation" is important.

b. Formation of the Treatment Relationship

The court in *Hurley v. Eddingfield, supra* absolved Dr. Eddingfield despite his having been the deceased's "family doctor." This reflects the general rule that an established custom of past treatment does not oblige a doctor to treat a patient's future illnesses; doctor/patient relationships are specific to a "spell of illness" and must be established, or renewed, accordingly.

Within a given "spell," however, the law often requires very slight involvement before finding that a treatment relationship between patient and doctor (or hospital) has been formed. A patient's description of symptoms over the phone followed by a physician's brief instructions, a telephone call to a physician's office for the purpose of initiating treatment, or scheduling an appointment to treat a particular medical problem have all sufficed

to support a factfinder's inference that a doctor or hospital had undertaken to provide care. While little is generally required, the decisions are not uniform: courts have also found that no relationship arose where the call to a physician's office to schedule an appointment did not itself seek or generate medical advice, and that where a patient interpreted the physician's response to her telephone contact as a refusal to undertake care, the requisite "consensual" characteristic of the relationship was missing—irrespective of the objective content of their communication.

Physicians' informal "curbside" consultations with colleagues normally will not establish a relationship between the patient and the consultee-physician.

80

Courts fear implying a treatment relationship out of limited, routine consultative contacts (of which the patient, incidentally, is often unaware) would chill a useful medical practice, to the detriment of patients and physicians alike. Of course, more formal physician referrals likely *will* result in legal recognition of the treatment relationship.

Finally, no treatment relationship customarily arises where physicians examine patients for the benefit of third parties. Thus physicians conducting physical exams for insurance eligibility or for employment-related purposes generally are not held liable to the examinee for failure to treat, or for other medical errors or nondisclosures. Exceptions have arisen, however. In the employment context, courts have implied a *limited* relationship, imposing a duty that extends only to disclosure of any test results that "pose an imminent danger to the examinee's physical or mental well-being," or have implied the relationship where the

physician affirmatively undertook treatment or gave advice. At least one case held that an employer *itself* (in contrast to the examining physician) may be liable to the examinee for negligent failure to disclose a serious medical problem discovered in a pre-employment exam. Dornak v. Lafayette General Hospital, 399 So.2d 168 (La. 1981). In contrast, another case found that a life insurer has *no* duty to disclose positive HIV test results to a policy applicant. Deramus v. Jackson Nat. Life Ins. Co., 92 F.3d 274 (5th Cir. 1996). However, the court, seemingly unaware of the "no duty" rule usually applied in such situations, suggested in dictum that if a physician (rather than

the company) had been "directly involved," the court might find a duty to disclose such information, based on patients' expectations, professional ethics, and physician expertise in health matters. Of course, to the extent these policy rationales are persuasive, they undercut the no-duty rule that normally applies in these cases.

2. HOSPITALS

It is sometimes stated that the "no duty" rule documented above for physicians applies with equal force to hospitals. While this *might* have been true at one time (the older case law seems to say so, but it is not without ambiguity), this general "no duty" rule unquestionably is not the law now. Hospitals and other health care institutions, in contrast with physicians, operate under numerous sources of law (both statutory and court created) that prohibit the arbitrary refusal to admit patients.

Before proceeding to an exploration of those developments, it is important to understand that uninsured patients traditionally have relied on free care rendered by public or private hospitals. Most

larger cities maintain a municipal hospital that is obliged to treat all patients regardless of the ability to pay, and many smaller localities historically provided for the uninsured by compensating private hospitals for treating the poor. This local largesse has become overtaxed, however, and so many municipalities have greatly scaled back or eliminated their support for medically indigent residents. Private hospitals have a long tradition of caring for

the poor, but their capacity for charity care has been stretched thin by reimbursement constraints that eliminate the ample revenues previously received from insured patients.

As a consequence of these various social forces, some private hospitals have turned desperately ill patients away from their emergency rooms, usually by transferring them to public municipal facilities. This practice of "patient dumping" has led both to litigation over a private hospital's obligation to render emergency care to indigent patients and to a new federal law addressing the practice (sec. A.2.b.). Because of this history and the persistence of access disparities, the following two subsections, while exploring hospital treatment duties generally, have particular importance for access to care by the indigent.

a. The General Duty to Provide Care

Wilmington General Hospital v. Manlove, 174 A.2d 135 (Del.1961) is the seminal decision that finds in the common law a duty on the part of hospitals to act reasonably in their patient selection decisions. *Manlove* involved a hospital emergency room that refused to treat a severely ill infant because he was under the care of another physician who was not a member of the hospital's medical staff. As a result, the infant died. By analogy to the tort of

negligent termination of gratuitous services, the court reasoned that in cases of "unmistakable emergency," a hospital that maintains an emergency room which by "established custom" has been open can properly

be held responsible for refusing to treat a patient whose condition "worsens" as a consequence of time lost pursuing the unforthcoming treatment. Detrimental reliance is thus at the core of the case. Accordingly, *Manlove* applies only to emergency care, and even then its scope is rather limited. (Section A.2.b.1, *infra*).

A potentially more powerful and sweeping common law theory—one that would cover *all* forms of hospital treatment—asserts that private hospitals owe duties to the public at large on the ground that they are "quasi-public," by virtue of the importance of their services, the funding they receive from public sources, their licensure, and their tendency to enjoy monopoly status in a community. Chapter 3.B.4 explores in some detail the notion, accepted in some states, that such "quasi-public" status has importance in the context of *physicians* seeking access to a hospital; it is even more to the point with respect to *patient* access. The *Manlove* court rejected this view, however, and it has not in fact been widely adopted, although a few courts have been receptive to it. *See, e.g.*, Leach v. Drummond Med. Group, 192 Cal.Rptr. 650 (App.1983) (reasoning applied to the only physician group practice in town).

b. Access to Emergency Care

Notwithstanding the absence of a general duty to rescue, in certain areas the law has been slowly (perhaps even ambivalently)

but perceptibly responsive to the moral challenge of taking action to reduce acute, visible and avoidable suffering. One of

the areas in which this trend may be at work is in the legal recognition of a hospital duty to care for emergency patients irrespective of their ability to pay.

(1) Common Law and Statutory Rights

Manlove was the first case to fashion a theory of relief for patients denied hospital emergency care. Its principal impact has been in securing access to emergency care by uninsured patients. In addition, about half the states have laws expressly requiring hospitals to treat emergency patients without regard to their ability to pay. Federal law imposes the same duty on hospitals that maintain charitable tax exemption (discussed in Chapter 5.D).

These established legal protections are limited, though, by their narrow definitions of what constitutes an emergency and of the extent of treatment required in an emergency. For instance, the *Manlove* theory, applied in a number of states, addresses refusals to treat only in cases of "unmistakable" emergency, only where the patient's condition worsens due to the delay in finding an alternative source of care, and only where the delay is caused by reliance on an ER's open-door custom. Many of the state statutes define an emergency as a situation requiring immediate treatment in order to prevent loss of life or limb—which can exclude a broad range of serious, albeit less extreme, medical conditions.

Two Arizona cases point the way toward a more expansive duty to provide emergency care.

Eschewing *Manlove*'s reliance-based approach, the Arizona Supreme Court has implied a sweeping duty "to provide emergency care to *all* persons presenting themselves for such aid" (emph. in original), ostensibly based upon the state regulatory requirement that all general hospitals maintain emergency facilities as a condition of licensure. Guerrero v. Copper Queen Hospital, 537 P.2d 1329 (Ariz.1975). A decade later the court relied on JCAHO standards prohibiting discrimination based on the "source of payment," incorporated by reference into the state's hospital licensing statute, to conclude that hospitals may never transfer emergency patients for economic reasons. Thompson v. Sun City Community Hospital, 688 P.2d 605 (Ariz.1984).

Guerrero and *Thompson* are important for two reasons. First, they are best understood as based in common law public policy—essentially, an emergency room application of the "quasi-public status" theory discussed above—rather than on idiosyncrasies of state regulatory law, and are thus of general rather than parochial interest. Second, they allow courts to redefine the nature of an emergency and the extent of the treatment required. Hospitals are obliged to treat any patient with a "need for immediate attention" and to provide such patients all care that is "medically indicated." The duty to treat thus encompasses far more than care necessary to prevent the patient's condition from deteriorating: "The relevant inquir[y] . . . d[oes] not relate to 'stabilization' and 'transferability,' but rather to the nature and duration of the emergency." 688 P.2d at 611. This broadened theory has not been

adopted by other states, perhaps due to the subsequent enactment of, and widespread reliance on, EMTALA—notwithstanding EMTALA's own arguable limitations in this connection.

(2) The Emergency Medical Treatment and Active Labor Act

The Emergency Medical Treatment and Active Labor Act of 1985 ("EMTALA"), 42 U.S.C. § 1395dd, has become the single most important legal tool governing access to emergency care, due principally to its uniform national applicability and its remedies. Hospitals that receive Medicare payment must comply with EMTALA's terms for *all* their patients. EMTALA creates a private right of action for damages for violation of its terms by such hospitals, though there is no comparable action against physicians. It also authorizes civil money penalties up to $50,000 for negligent noncompliance by both hospitals *and* physicians.

EMTALA was enacted in the belief that state law was too weak to prevent the widespread "dumping" of indigent and uninsured patients. Its protections, however, go further: they are triggered by the refusal to properly examine or treat "any individual" who comes to a hospital emergency department seeking care, irrespective of the person's eligibility for Medicare or whether he can pay for care. EMTALA requires, first, that the hospital provide for an "appropriate medical screening examination within the capability of the hospital's emergency department," to determine whether there is a

medical emergency. If so, treatment must then be provided to the point of "stabilization." Specific analogous provisions also apply to women in labor.

(a) Screening

A moment's reflection on the statutory language just quoted suggests that a diagnostic screening might be "[in]appropriate" in

varying ways, and for different reasons. Uncertainty has thus arisen over just what hospital conduct the statutory phrase reaches, as well as what standard of performance it imposes. Initially, there was debate over whether EMTALA requires or assumes an improper *motive* for refusing treatment, based on factors such as economics, demographics, or personal characteristics. This has been criticized on the grounds that there is no statutory support for such a reading, that it is so inclusive as to be virtually without limit (and therefore meaning), and that it is nonetheless sufficiently difficult to prove that it would defeat virtually all EMTALA claims. In Roberts v. Galen of Virginia, 525 U.S. 249 (1999), the Supreme Court put this issue to rest, holding that EMTALA's stabilization requirement (and probably also its screening requirement) applies regardless of motive

Instead of a subjective motivation test, many courts have applied a more objective test that looks to whether the patient was harmed by "disparate" or "non-uniform" treatment—some purposeful *variation* from the medical practices that the hospital would otherwise apply to similarly-situated persons.

This approach focuses solely on whether a hospital complied with its own standard procedures, and not whether those practices are themselves reasonable. Courts have adopted this "disparate treatment" analysis largely in an effort to avoid making EMTALA a federal malpractice law that would displace ordinary state-law negligence claims, a result which they believe Congress did not intend. Vickers v. Nash General Hosp., 78 F.3d 139, 142–43 (4th Cir. 1996) (reviewing cases).

However, a type of malpractice law seems inevitably to arise from the judicial requirement of uniform treatment. Uniformity presupposes the hospital has a standard practice against which to measure the alleged variation, and indeed this is often true. Hospitals are compelled by many forces—state licensure laws, their own governing boards and by-laws, JCAHO standards, the threat of malpractice liability—to adopt policies and procedures that are normative, likely embracing accepted standards of professional competence by institutions of comparable size, nature and circumstances, rather than merely allowing idiosyncratic and perhaps sub-standard conditions or practices to take hold unheeded. This is the stuff of malpractice law.[1] Courts will not likely accept a hospital's defense that it *has* no standards against which divergence could

be measured, for to do so would encourage such conduct; instead they will impose the relevant tort-law standard of care on the institution. Power v. Arlington Hospital Ass'n., 42 F.3d 851, 858 (4th Cir. 1994). Thus in purporting to require only "equal" or *consistent* care by a given hospital, the disparate treatment test actually imposes a substantive requirement of *nonnegligent* care.

Where an institution deviates from its (nonnegligent) standards through medical error or omission, many courts nevertheless refuse to apply EMTALA, out of a continuing reluctance to federalize all emergency room tort law. Instead, they require a purposeful deviation. Thus where a physician failed to detect chest sounds indicating a broken sternum and rib which would normally prompt a diagnostic X-ray, the court ruled that the failure to X-ray was simple negligence and *not* "disparate treatment" sufficient to state

an EMTALA claim. Summers v. Baptist Medical Center Arkadelphia, 91 F.3d 1132 (8th Cir. 1996) (*en banc*). Compare this approach to the *Power* case, where the court characterized the hospital's failure to give the plaintiff a diagnostic blood test—the allegedly negligent error *itself*—as "disparate treatment" sufficient to constitute an EMTALA claim. Although the court struggled to sustain a formal distinction between EMTALA and malpractice law by observing that only negligent *omission* of a diagnostic test and not negligent *interpretation* or *performance* would support claims under both theories, this decision virtually conflates EMTALA and medical negligence in all cases of omitted emergency treatment.

90

(b) Treatment and Stabilization

If the required screening reveals an "emergency medical condition," the hospital must undertake treatment. EMTALA's definition of "emergency" is a condition reasonably likely, without "immediate" treatment, to create "serious jeopardy" to the person's health. This definition is at least as inclusive as many of those found in state common law and state statutes, and the range of conditions it covers is thus reasonably broad.

The more problematic issue is how far treatment must proceed under the statutory mandate to "stabilize" the condition. Stabilization is defined as a level of treatment likely to prevent "material deterioration" of the condition during transfer. (Transfers of unstabilized patients are permitted in limited, specified circumstances involving a written request or expected medical benefit). The facts of a pre-EMTALA case, Joyner v. Alton Ochsner Medical Foundation, 230 So.2d 913 (La.App.1970), are instructive. An auto accident victim came to the emergency room

of a private hospital with "multiple deep facial lacerations, a possible head injury, traumatic damage to the teeth and multiple bruises and contusions of the body, resulting in considerable loss of blood." The hospital merely bandaged him, took X-rays, monitored for shock and administered I.V. fluids to stabilize his blood pressure before transferring him to a Veteran's Administration hospital for further treatment. This course of action is probably entirely consistent with EMTALA; thus in many situations, EMTALA actually may not

require more extensive care than was due under state common and statutory law. Ironically, the expansive treatment obligations in a few common law precedents (*see Guerrero* and *Sun City*, *supra*) may actually *exceed* the "stabilization" requirement of EMTALA. Given EMTALA's dominance of the field, however, their further application seems unlikely, even though EMTALA does not actually preempt such common law claims.

(c) "Preventive" Dumping

EMTALA requires screening and stabilization of anyone who "comes to" an emergency department. Under this language courts have rejected EMTALA claims by patients who do not, literally, show up at the hospital. *E.g.,* Johnson v. University of Chicago Hospitals, 982 F.2d 230 (7th Cir. 1992) (dismissing EMTALA claim where hospital telemetry staff directed paramedics treating child in full cardiac arrest to another hospital); Miller v. Medical Center of Southwest Louisiana, 22 F.3d 626 (5th Cir. 1994) (patient did not "come to" a hospital that refused, by telephone, to take him as a transfer on economic grounds). By regulation, the Secretary of Health and Human Services has confirmed that "comes to" requires

physical presence on hospital property; however, a hospital's own ambulance is *deemed* hospital property, and, arguably, non-hospital ambulances cannot be re-routed except for reasons of lack of hospital capacity or staff.

EMTALA has engendered debate. Some criticize it for responding to a problem that was never as

widespread as claimed and for imposing an awkward and ill-drafted solution. Others believe that EMTALA has been effective; still others that patient dumping persists at unacceptable levels, notwithstanding EMTALA. Under any view, however, EMTALA has become central to the law of access to emergency medical care.

3. DOCTORS WITHIN HEALTH CARE ORGANIZATIONS

How can hospitals, dependent on doctors to deliver care, comply with their institutional duty to treat (under the various legal theories explored above) if *Hurley* leaves physicians free to refuse patients? One solution is regulatory: since EMTALA was enacted, physicians are no longer completely free to refuse emergency patients with impunity, because they may face civil fines for negligent noncompliance with EMTALA's terms. A second solution (pre-EMTALA) is contractual, provided by Hiser v. Randolph, 617 P.2d 774 (Ariz.App.1980). Hospitals may require as a condition of medical staff membership that physicians assist in treating emergency and indigent patients. If physicians accept this condition by joining the medical staff or working in the emergency room, then this contractual obligation may extend to the patient as a third-party beneficiary.

A similar solution applies to managed care, where a health plan may contractually bind participating physicians to see individuals it has a contractual duty to treat. In Hand v. Tavera, 864 S.W.2d 678 (Tex. App. 1993), the court relied on the applicable

contracts (characteristic of health plans) to find a treatment relationship with an on-call plan physician. The physician refused to authorize the patient's admission to the plan's hospital based on the symptoms and history conveyed by a telephone consult from the ER. As a consequence, the patient had a stroke at home. The court reasoned that the enrollee paid premiums to the plan to purchase medical care in advance of need; the plan arranged to meet its obligation to provide care by paying physicians; those physicians, in return, agreed to treat the plan's members. The *identity* of the physician who happened to be on call for emergency admissions was immaterial: the plan brought the patient and physician together "just as surely as though they had met directly and entered the physician-patient relationship." 864 S.W.2d at 679.

4. WRONGFUL DENIALS: ANTIDISCRIMINATION LAW AND REFUSAL TO TREAT

As discussed thus far, physicians (and, to a considerably lesser extent, hospitals) enjoy substantial legal discretion to refuse patients for "good" reasons, "bad" reasons, or no stated reason at all. In a limited number of areas, which are the subject of this section, federal law specifically disapproves certain bases for treatment refusals. In addition, states often have counterpart regulatory laws, generally applicable to "public accommodations" (which covers hospitals but often not medical offices). The

JCAHO's accreditation standards also prohibit discriminatory practices by

hospitals on the basis of race and other characteristics, including source of payment, and the federal charitable tax exemption for hospitals carries with it certain obligations to provide care on a nondiscriminatory basis to paying patients.

a. Title VI: Race and Ethnicity

Title VI of the federal civil rights law, enacted in 1964, prohibits any "program or activity receiving federal financial assistance" from discriminating against, excluding, or denying benefits to individuals on the grounds of race, color, or national origin. 42 U.S.C. § 2000d. Thus, overt discrimination by health care institutions participating in the federal Medicare or Medicaid programs (which were enacted in 1965) or receiving other financial support is barred. Despite Title VI, subtler forms of racial (as well as gender) discrimination by health care providers doubtless persist. Title VI has been invoked in a few cases challenging decisions to relocate or to close hospitals serving predominantly minority populations. For much of its history, however, Title VI did not apply to physicians. Section 1557 of the Affordable Care Act closed this gap, and expanded the bases of prohibited discrimination to include sex/gender.

b. Disability Discrimination

Two closely related laws, both of which apply to a wide range of activities beyond health care, are of rapidly growing importance in this field: Section 504 of the Rehabilitation Act of 1973 (29 U.S.C. § 794),

and the Americans with Disabilities Act of 1990 (42 U.S.C. § 12101). HIV/AIDS first drew attention to the application of disability law to health care and HIV cases continue to arise and receive coverage, but the application of disability discrimination law to health care is considerably broader.

The most obvious difference between the two laws is the reach of their regulation. Section 504 applies to *federally funded* "programs and activities" (which includes hospitals that receive Medicare reimbursement, but probably not doctors). The ADA, by contrast, reaches various entities irrespective of whether they receive federal financial assistance, including state and local governments (Title II) and public accommodations (Title III). Of greatest importance here, the latter include the "professional office of a health care provider, hospital, or other service establishment," so doctors' offices are covered. While there is more case law developed under § 504 because it has been in effect for so much longer, the greater reach of the ADA suggests that it will ultimately supersede § 504 in importance.

(1) Protected Class

The first question in most disability discrimination cases is whether the individual falls within the protected class. Section 504 protects a "handicapped individual," defined as someone with a "physical or mental impairment which substantially limits one or more of such person's major life activities," or someone with either a "record of," or who is "regarded as having," such an impairment. The ADA's

definition of "disability," except for the choice of the operative word, is almost verbatim. 42 U.S.C. § 12102(2).

These terms are quite broad, reflecting Congress' intent to protect people against discrimination arising not only from prejudice but also from fear and myth. School Bd. of Nassau County v. Arline, 480 U.S. 273, 279, 284 (1987). Echoing those policies, the Supreme Court held that non-symptomatic HIV infection constitutes a disability under the ADA. Bragdon v. Abbott, 524 U.S. 624 (1998). In *Bragdon* an HIV-positive patient alleged that her dentist violated the ADA when he refused to drill a cavity for her in his office (offering, instead, to do so at a hospital, though there was no evidence the hospital would be safer or even that he had privileges to practice there). The Court held that non-symptomatic HIV infection constitutes (in the statutory terms) a "physical impairment" from the moment of infection onward, and that, by interfering with the plaintiff's reproductive capacity, the infection "affected a major life activity" because of the centrality to life of reproduction and sexual relations. The open-ended nature of this "major life activity" category is suggested by the Court's intimation that other plaintiffs might persuasively assert that HIV impacts *other* life activities, as well. Finally, the Court concluded that HIV infection was a "substantial limit" on the plaintiff's reproductive activity, noting that her status would impose significant risks of infection on male sexual partners (20–25%), and on any child during gestation and childbirth (8–25%). The Court emphasized that this

third requirement is met "even if the difficulties [generated by the disability for the life activity in question] are not insurmountable."

Bragdon bolsters a broad construction of "disability." For instance, recognition that reproduction is a "major life activity" suggests that a health insurer's exclusion of infertility treatments might be challenged as disability discrimination. But, *Bragdon* may well be the apex of expansive interpretation of disability law. The following year, the Court ruled that pilots with imperfect eyesight do not have a statutory disability, even though this kept them from flying for commercial airlines, because their impairment was easily enough corrected with normal prescription glasses. The fact that they still did not qualify them for commercial flights did not meet the "regarded as" disabled standard because the airline was willing to hire them for non-piloting jobs. Sutton v. United Airlines, 527 U.S. 471 (1999).

(2) Core Provisions

Finding a handicap or disability is only the first inquiry. Section 504 prohibits regulated programs or activities from excluding, denying benefits to, or discriminating against any "otherwise qualified handicapped individual . . . solely by reason of his handicap." 29 U.S.C. § 794. "Otherwise qualified" means able to meet program requirements "in spite of" the handicap, as established through an individualized, factually-specific inquiry. Even where a person cannot initially meet all program or activity requirements, he or she may nonetheless be

"otherwise qualified" if the sponsor of the program or activity can make "reasonable accommodation"—i.e., take steps, short of incurring "undue financial and administrative burdens" or making "a fundamental alteration in the nature of the program" (*Arline,* supra), that would enable the person to meet the requirements, in

which case the sponsor must do so. The ADA rules, though not identical, are similar. One commentator has concisely summarized the operative core of the two laws:

> The superficially distinct requirements that challenged conduct both disfavor "qualified" disabled applicant and also result in discrimination "on the basis of" disability typically collapse into a single inquiry. As the U.S. Supreme Court observed in *Alexander v. Choate*, [469 U.S. 287, 299 n. 19 (1985)], "the question of who is 'otherwise qualified' and what constitutes improper 'discrimination' . . . [are] two sides of a single coin." A person who lacks legitimate qualifications has not been impermissibly discriminated against. Under both statutes, a person is "qualified" to receive services such as health care if, with reasonable modifications, she is able to meet a program's "essential" or "necessary" eligibility requirements.

Philip G. Peters, Jr., *Health Care Rationing and Disability Rights*, 70 Ind.L.J. 491, 507 (1995).

A key difficulty in disability law is determining whether a person is "otherwise qualified" for the benefit or service. Classical applications of this

standard arise in cases involving access to education and employment, in which the analysis has two salient characteristics: (1) it impliedly assumes that the benefit or service is generally available to a qualified class of people, under terms established by its sponsor; and (2) because the disability is not the *reason* for which the person seeks the benefit or service, it is coherent to ask whether the person can meet the eligibility terms *notwithstanding*

(or "in spite of") his disability, with any needed reasonable accommodation.

In health care, this analysis applies logically enough where an individual is seeking access to care for a problem that is *un*related to his disability, as where a physician who is treating a patient for an ear infection refuses to perform medically indicated surgery after he learns the patient is HIV positive. The benefit (ear surgery) is generally available on certain terms to those who need it; and, since the patient is seeking care for a condition other than his disability (HIV status), it is coherent to ask whether, with reasonable accommodation (here, relating to his infectiousness and his immunocompromise), he qualifies for the benefit "in spite of" that disability.

But in cases in which it is precisely the disability that *gives rise to* the need for access to health care, this is not a very coherent approach to "otherwise qualified." In Wagner v. Fair Acres Geriatric Center, 49 F.3d 1002 (3d Cir. 1995), for example, a nursing home with many Alzheimer's patients denied admission to an agitated, violent 65-year-old Alzheimer's patient on the ground that the home was

inadequately staffed to care for such patients. If one takes at face value the service the home *actually* offered (care for *non-violent* Alzheimer's patients), then the service that the plaintiff *sought* (care for *violent* patients) was simply not one that the home made generally available to a qualified class; viewed this way, the first element of the usual "otherwise qualified" analysis could not be met. It was only by judicially redefining the eligible class—determining that violent patients *should* be admissible to the home

(a conclusion buttressed by evidence that the plaintiff could, in fact, be accommodated without unreasonable burden on the home)—that the court could uphold the jury's finding that the plaintiff was "otherwise qualified." Observe, though, that under this expanded view of eligibility, the *second* element of the traditional analysis now becomes incoherent: it makes no sense to ask, "in spite of the fact that she has Alzheimer's disease, is a person eligible for care at a nursing home that takes Alzheimer's patients of all kinds?"

A few courts have found a more meaningful approach in these cases. They preclude a health care provider from using disability *alone* as the basis for withholding medical benefits. A person is "otherwise qualified" for a particular medical benefit if "there is no factor [other than a bona fide medical reason] apart from the mere existence of disability that renders the participant unqualified for the benefit." Woolfolk v. Duncan, 872 F.Supp. 1381, 1389–90 (E.D. Pa. 1995) (seeking care for, and alleging discrimination based on, HIV status). This would allow a physician to refuse care based on the claim

that the physicians felt under-qualified to treat the patient's particular condition. In Lesley v. Chie, 250 F.3d 47 (1st Cir. 2001), the court took a deferential attitude toward this claim, holding that an obstetrician could refer an HIV-infected pregnant woman to another hospital as long as this asserted reason was not "arbitrary and capricious," a mere "pretext for some discriminatory motive," or "devoid of any reasonable medical support."

5. OTHER BASES FOR A DUTY TO TREAT

a. Constitutional Rights of Access

It is clear that there is no federal constitutional obligation for government to *fund* health care. The cases involving abortion funding have firmly established that "the Constitution imposes no obligation on the States to pay . . . any of the medical expenses of indigents." Maher v. Roe, 432 U.S. 464, 469 (1977). The due process clause traditionally has been seen as protecting individuals *from* improper government interference ("negative" liberties), rather than generating entitlements *to* state-conferred assistance or benefits ("positive" liberties). E.g., Wideman v. Shallowford Community Hospital, 826 F.2d 1030 (11th Cir.1987). *See also* Deshaney v. Winnebago County Department of Social Services, 489 U.S. 189 (1989) (echoing the foregoing).

Violations of negative liberties might be argued in many contexts of health care regulation. One court found a generalized constitutional right to be free of poorly justified state restrictions on medical

decisionmaking. Andrews v. Ballard, 498 F.Supp. 1038 (S.D.Tex.1980) struck down a law that allowed only licensed physicians to practice acupuncture as an infringement of patients' right to "obtain or reject medical treatment," which the court found was encompassed by the right to privacy identified in *Roe v. Wade.* Most courts, however, require only a rational justification and, in any case, they view protecting health as a compelling state interest, so these types of argument rarely succeed. For instance, courts have upheld state bans on alternative cancer therapy that is probably harmless but thought to be ineffective. *Cf.* United States v. Rutherford, 442 U.S. 544 (1979).

Even though there is no general constitutional right to health care (as there is in some European countries), "when a State does decide to alleviate some of the hardships of poverty by providing medical care, the manner in which it dispenses benefits is subject to constitutional limitations" imposed by the due process and equal protection clauses. *Maher, supra,* 432 U.S. at 470. Thus, a publicly-funded hospital cannot arbitrarily discriminate in the patients it treats or in the services it provides. *See* Memorial Hospital v. Maricopa County, 415 U.S. 250 (1974) (unconstitutional to refuse county health services to temporary residents).

Another exception to the principle that only "negative liberties" generally enjoy constitutional protection arises when the state has "control" over an individual. Notable litigation has arisen over the

treatment rights of institutionalized persons, particularly the mentally ill. The federal courts have held that it is unconstitutional to confine patients involuntarily for the purpose of treatment and then provide no treatment. O'Connor v. Donaldson, 422 U.S. 563 (1975). The most remarkable instance is a federal judge who took direct charge of the administration of Alabama's state medical hospital because of its persistent failure to provide any meaningful form of treatment. Wyatt v. Stickney, 325 F.Supp. 781 (M.D.Ala.1971); 344 F.Supp. 373 (M.D.Ala.1972). Prisoners also gain certain rights to health care by virtue of their confinement, under the 8th Amendment's prohibition of cruel and unusual punishment. *See* Estelle v. Gamble, 429 U.S. 97 (1976).

b. Patients' Rights

A final source of patients' access rights are so-called Patients' Bills of Rights. In order to comply with JCAHO accreditation

standards, hospitals must adopt statements that cover a variety matters related to patient care, including access to care and patient dignity and confidentiality. Although these statements are issued voluntarily, courts probably would give them binding legal force as forming part of the hospital's contractual relationship with its patients. Moreover, in some states these statements are not voluntary. Minnesota is one among several jurisdictions that specify a mandatory bill of rights. On the federal level, a detailed set of protections applies to patients in nursing homes that receive

Medicare (42 U.S.C. § 1395i–3) or Medicaid (42 U.S.C. § 1396r) funding.

The rapid growth of managed care has been accompanied by concerns that health plans are restricting subscribers' access to care by a variety of mechanisms including limiting referrals, restricting coverage, and forcing physicians to withhold information about treatments not offered or covered by the plan (so-called "gag" clauses). As a consequence there has been significant state legislative activity enacting new patient rights of access. These laws typically require a broader choice of physicians, provide mechanisms to appeal negative coverage decisions and prohibit restrictions on physician communication with patients. Some, but not all, of these provisions were also included in the federal Patient Protection and Affordable Care Act of 2010.

6. TERMINATING THE TREATMENT RELATIONSHIP

We began this chapter with the observation that professional duties arise upon the formation of a treatment relationship and continue until it is properly terminated. Patient "abandonment" is the term applied to an improper termination of treatment that is

intentional, in contrast with termination that is due to a mistake in medical judgment. The latter is a matter for ordinary malpractice law, but this distinction is frequently confused.

105

Where a treatment relationship exists, the law of abandonment requires that the physician (or hospital) provide all necessary care unless the relationship is terminated (1) by the patient or (2) by the provider, after giving the patient proper notice and an opportunity to secure an alternate source of care. Abandonment law is thus much more forgiving than is usually recognized. The only explicit restraint on a doctor's (or hospital's) freedom to abandon a patient is the *procedural* one of notice. As classically conceived, there is no real *substantive* content to abandonment law because the law does not scrutinize the *reasons* for abandonment: so far as abandonment law is concerned, a doctor may, with proper notice, stop treatment because he wants to retire, or go on vacation, or simply because he dislikes the patient.

To ensure that there is no question about the adequacy of notice to the patient and opportunity to obtain substitute care, prudent physicians as a practice usually take affirmative steps themselves to arrange for substitute care. A vacationing doctor will usually have an associate cover her cases and a retiring doctor will ordinarily tell patients that a designated physician has agreed to take her cases. Prudent hospitals, likewise, will never simply discharge an ill patient, even after ample notice; instead, they will locate an alternative facility to which a patient can be transferred.

These pragmatic accommodations have created a degree of uncertainty in abandonment law. Because this body of law is based on an implied contractual

undertaking and on notions of fiduciary responsibility, its precise limits are not firmly set. Consequently, the case law in different states offers conflicting indications of whether simple notice of treatment termination is sufficient, or whether instead the law requires health care providers to arrange for a substitute source of care.

This point becomes a critical issue in the modern context where doctors and hospitals face increasingly severe constraints in health care reimbursement. A provider might seek to terminate care because the patient's insurance runs out or won't cover the treatment. If patients are given "notice," and perhaps appeal rights, will this suffice under the usual procedural requirements of abandonment doctrine—or might a creative and sympathetic plaintiff persuade a court to read a substantive element into the doctrine, and prohibit the termination of treatment based on inability to pay?

Three cases shed conflicting but ambiguous light on the legality of "economic abandonment." In Ricks v. Budge, 64 P.2d 208 (Utah 1937), the court allowed the plaintiff to maintain an action alleging the following facts: After ordering Mr. Ricks to the hospital for a seriously infected hand, Dr. Budge refused treatment and walked out because Mr. Ricks would not immediately catch up on his past due accounts. This decision is frequently cited by commentators for the proposition that it is illegal to abandon a patient who cannot pay. However, these facts to not support a general prohibition of economic abandonment. Instead, the holding is perfectly

consistent with purely "procedural" abandonment law, which only requires the doctor "to give the patient sufficient notice so the patient can procure other medical attention if he desires," *id.,* and prohibits patient abandonment only at a critical stage in the course of treatment.

Additional support for this view is suggested in a leading abandonment decision, Payton v. Weaver, 182 Cal.Rptr. 225 (1982). There, the court allowed a physician to stop treating an uncooperative patient despite the unavailability of any substitute care. This case concerned a renal dialysis patient who "frequently appear[ed] for treatment late or at unscheduled times in a drugged or alcoholic condition, used profane and vulgar language, and on occasion engaged in disruptive behavior, such as . . . cursing staff members with obscenities." Although Dr. Weaver's attempts to find alternative treatment centers were unsuccessful, the court held that he "gave sufficient notice of [his intent to cease treatment] and discharged all his obligations." It is impossible to determine, though, whether *Payton v. Weaver* invokes a purely procedural abandonment rule because its compelling facts might also provide substantive justification for the decision to discontinue treatment.

Finally, in Muse v. Charter Hosp. Winston-Salem, Inc., 452 S.E.2d 589 (N.C. App. 1995), aff'd mem., 464 S.E.2d 44 (N.C. 1995), the court ruled that a psychiatric hospital illegally interfered with the physician's medical judgment when it encouraged the discharge of an adolescent patient whose insurance

ran out, which led to the patient's suicide three weeks later. Although the opinion never mentions the abandonment doctrine, it is based on the hospital's alleged "policy or practice" of

discharging patients when their insurance runs out. Even so, the decision is subject to the same uncertainty as *Ricks* and *Payton*: we don't know whether the hospital is liable because the discharge policy is per se wrong or because of the procedure it followed in failing to sufficiently notify the parents of their son's fragile condition so they would be sure to find alternative care. Even with such notice, however, a patient who is discharged for financial reasons is likely to have a difficult time finding alternative care.

B. THE LEGAL CONTENT OF THE TREATMENT RELATIONSHIP

The balance of this chapter addresses the legal doctrines that arise once a treatment relationship is formed: confidentiality, informed consent (including conflicts of interest), and contractual modification of treatment obligations. The physician's duties under each of these doctrines are shaped by the fiduciary nature of that relationship.

1. THE FIDUCIARY CORE OF THE TREATMENT RELATIONSHIP

"Fiduciary" relationships exist in a number of legal and social realms. Fiduciary duties arise as heightened aspects of general tort and contract law rather than through a separate branch of legal doctrine. Fiduciaries must meet high standards of

loyalty, diligence, and solicitude in carrying out their legal obligations. Black's Law Dictionary is instructive: "scrupulous good faith and candor"; the responsibility to act "primarily for another's benefit" in connection with a duty undertaken; the subordination of one's personal interests to that of another; a

relationship founded on the "trust or confidence" one person reposes in "the integrity and fidelity" of another who, as a result, can exercise "domination and influence"; the "highest standard of duty imposed by law"; trusteeship. A slightly weaker form of fiduciary status exists in what is called a "confidential relation," which arises when "on the one side there is an overmastering influence, or, on the other, weakness, dependence, or trust, justifiably reposed." Contrast these notions with normal arm's-length market relations.

In the treatment relationship, a patient, often dependent and diminished by illness, seeks care from a professional with a complex body of knowledge and skill essential to the patient's well-being and perhaps to life itself. He entrusts his care to the doctor, which often requires the sharing of intimate knowledge and deep invasions of physical and emotional privacy. These and other characteristics clearly bring it within the ambit of a fiduciary relation, and courts have consistently so held. "The patient's reliance upon the physician is a trust of the kind which traditionally has exacted obligations beyond those associated with arms-length transactions. His dependence upon the physician for information affecting his well-being, in terms of contemplated treatment, is well-nigh abject." Canterbury v.

Spence, 464 F.2d 772 (D.C. Cir. 1972). These characteristics form common underpinnings for the legal doctrines explored in the balance of this chapter.

2. CONFIDENTIALITY

Confidentiality is a core element of the fiduciary content of the treatment relationship. This section explores the values it protects,

the means by which the law safeguards it, and the tensions it creates in competition with other values.

a. The Duty to the Patient

So basic is the expectation of confidentiality in medical treatment that patients and health care providers alike generally assume it will be honored, without ever specifically discussing it (though clinical psychologists routinely announce at the outset of therapy that certain information may require its breach). Confidentiality is a foundational principle of medical ethics, recognized in both the Hippocratic Oath ("Whatsoever I shall see or hear in the course of my profession . . . if it be what should not be published abroad, I will never divulge, holding such things to be holy secrets") and the AMA Principles of Medical Ethics ("A physician shall . . . safeguard patient confidences within the constraints of the law").

The main rationale for honoring medical confidentiality is utilitarian. Confidentiality is thought to encourage individuals to seek medical care who might otherwise avoid doing so out of

shame, embarrassment or fear of disclosure of their affliction; this benefits the sick, the currently well (in providing reassurance about similar protection for the treatment of their own future illness), and society as whole in the form of improved public health. The philosopher Sissela Bok identifies three other reasons for honoring confidentiality: a recognition of the interest in autonomy over personal information; the legitimacy of not only *having* secrets, but of *sharing* them (i.e., respect for disclosure and relational intimacy); and the special obligation that arises from the act of having *promised* not to disclose. Sissela Bok, Secrets 119–24

(1982). For these reasons, the duty of confidentiality is protected through a variety of legal sources.

(1) Common Law Protections

Redress for providers' unauthorized disclosure of patient information has been sought using theories including infliction of emotional distress, malpractice, breach of a confidential relationship or of fiduciary duty, invasion of privacy, and breach of contract. *See* Humphers v. First Interstate Bank of Oregon, 696 P.2d 527 (1985) and cases discussed therein. This variety probably reflects uncertainty, in particular jurisdictions, as to what theory is best adapted and likeliest to prevail, but strategic considerations may also be at play, relating to whether expert testimony is needed (as in a malpractice claim), the availability of damages and the existence of damage caps, and comparative limitations periods. Where the claim is brought in tort, courts may require that an enforceable *duty* of

confidentiality be established by reference to sources of settled law or policy, and will look to the professional ethical standards noted above, as well as to the jurisdiction's various statutory provisions, for evidence of that duty and its limits.

(2) Statutory Protections

Many state licensure laws provide that a breach of patient confidence constitutes unprofessional conduct that will subject a physician to discipline or license revocation. Such laws vary as to whether the violation must be intentional or whether a merely negligent disclosure will suffice, and they generally acknowledge that other provisions of law may create exceptions.

In addition, most states have evidentiary rules which prohibit certain health care providers (generally doctors or psychotherapists) from disclosing patient confidences. These "privilege" rules generally apply only to testimony, and to the discovery and/or admissibility of records, in judicial proceedings. Accordingly, they confer only limited protection of patient confidentiality. Even in litigation the professionals to whom they apply is variable, *see, e.g.*, Buchanan v. Mayfield, 925 S.W.2d 135 (Tex. Ct. App. 1996) (privilege inapplicable to communications with dentist), they are easily waived by the patient, and they are subject to many exceptions. There is no physician-patient privilege in the Federal Rules of Evidence, although the Supreme Court has implied a psychotherapist-patient

privilege under Rule 501, Jaffee v. Redmond, 518 U.S. 1 (1996).

Disease (or subject)-specific statutes creating confidentiality duties have increased in number. Many states have HIV-protective statutes, enacted out of concern over discrimination and ostracism of HIV patients. Federal law imposes confidentiality requirements on records of patients in federally-assisted drug and alcohol treatment programs, 42 U.S.C. § 290dd–2. More than half the states have enacted legislation addressing the discriminatory use of genetic information by insurers or employers, and there is some federal regulatory law on the subject; there is a great deal of debate concerning the adequacy of protection of genetic information more generally. As with any protective law, the precise reach of these statutes is sometimes uncertain. E.g., Doe v. Marselle, 675 A.2d 835 (Conn. 1996) ("willful" disclosure of HIV information means knowing or intentional, as opposed to inadvertent, but does not require intent to cause harm to the patient).

The patchwork nature of these various protective laws, and the increasing proliferation and computerization of medical information, prompted the federalization of privacy law. A statute known as HIPAA, for Health Insurance Portability and Accountability Act, which mainly deals with insurance regulation, required the Secretary of Health and Human Services to issue a massive medical privacy regulation, in order to make it safer to store authority to store and transmit medical information electronically (ostensibly a cost-savings

measure). The federal privacy rule, issued in 2000 and 2002, 45 C.F.R. § 164.501, governs all individually identifiable "personal health information" maintained by "covered entities," which include insurers and institutional providers, rather than just physicians. These entities must (1) adopt internal procedures to protect the privacy of protected health information; (2) keep records showing they have notified patients of their privacy policies; and (3) train employees regarding privacy procedures.

The complex regulatory framework addresses a host of other issues, such as what constitutes safe storage of information and the responsibility of covered entities to ensure that business associates comply with privacy requirements. The federal rules does not require specific written consent for most standard uses of personal health information, such as those related to "treatment, payment, or health care operations." But patients must be given an opportunity to agree or object to other types of uses or disclosures, such as research, or even simply notifying family members. A violation of the regulations may result in a significant civil penalty or criminal liability or both, but the regulations do not create a private cause of action or remedy.

b. The Duty to Protect Third Parties and the Limits of Confidentiality

In some circumstances a health care provider owes a duty to parties *outside* the treatment relationship, arising from an important interest or value in

competition with the duty to the patient. Failure to meet this competing duty may result in civil or even criminal liability, and fulfilling it sometimes (but not inevitably) requires a breach of patient confidentiality. Broadly speaking, the applicable rules may be broken into three categories, two statutory and one common law.

(1) Statutory Duties to Report

The most common and time-honored exception is the required reporting of various communicable diseases to state public health authorities. The list of conditions varies somewhat from state to state but AIDS, or in many states HIV infection alone, is common. In a similar category fall reports of child and elder abuse, alcohol and drug abuse, and, in some states, uncontrolled epilepsy among licensed drivers. Where the reportable condition is not uniquely dependent on medical diagnostic expertise, the statutory duty may run to other health care providers, and sometimes laypersons, as well; in those instances the tension with a duty of confidentiality may be reduced or eliminated. The purpose of all such requirements is to enable authorities to protect identified individuals or the community at large.

These reporting statutes sometimes explicitly grant the provider immunity from liability to the patient for any resulting breach of confidence. Such immunity is readily implied in any event, where

the patient's condition unambiguously meets a clear reporting obligation. (Of course, notwithstanding

this immunity, as a practical matter the treatment relationship may well be fractured by the mandated disclosure). Under some statutes, *failure* to report can result in civil or criminal sanctions, and doctors who do not report might be found per se liable to anyone who is injured.

A second class of mandatory reporting exists for knife or gunshot wounds that appear to be non-accidental. Here the policy is generally to catch and punish wrongdoers, rather than to protect third parties against continuing or future harms. One might argue that this constitutes a less compelling basis for the breach of confidentiality. On the other hand, the instrumental *purpose* of protecting confidentiality—encouraging treatment—may be damaged to a lesser degree in the case of such violent injuries, since the victims (like those in any emergency) may be highly likely to obtain treatment out of necessity.

(2) Common Law Duty to Protect Third Parties

Even absent a statutory duty, a legal duty to protect third parties might arise through the common law whenever the patient's condition poses a significant risk or danger to others. Tarasoff v. Regents of University of California, 551 P.2d 334, 340 (Cal. 1976). Examples include patients with contagious diseases, violent psychiatric patients, and persons with medically-related driving impairments. It is often extremely difficult to resolve these situations under the common law. The existence and scope of the duty to third parties is often unclear;

moreover, in those cases where this duty competes with the obligation of confidentiality, the tension is likely to be very great and immediate, and the consequence to the treatment relationship of a breach of confidence quite destructive. It is often the case that a physician has no legally safe course available.

Note at the outset that, although these common law cases are frequently described as imposing a duty to *warn* third parties, the category is better conceptualized as a duty to *protect* third parties, since there may be steps other than direct warning that are necessary or sufficient to discharge the duty. Such steps may not require breaching confidentiality.

Broadly speaking, physicians are liable for harm to a third party if three conditions are met: (1) there is a known or reasonably foreseeable hazard, arising in some way from the physician's patient; which (2) places at risk one or more foreseeable (though not necessarily individually identifiable) third parties, of whom the plaintiff is one; and (3) the provider failed to take a reasonable course of protective action. The first condition may be thought of as giving rise to the *existence* of the duty; the second determines the party(ies) to whom the duty is *owed*; and the third defines the *scope* of the duty.

(a) Basis of the Duty: "Hazard" or "Special Relationship"

The familiar basic tort rule is that one has no duty to take protective action on another's behalf simply

because one recognizes (or should recognize) that the other is in avoidable jeopardy. Courts have historically recognized exceptions

for "special relationships" that give rise to a duty to control someone's conduct for the benefit of a third party (e.g., hospitals' relationships with mental inpatients). *See generally* Restatement (Third) of Torts § 40–42 (2005). In practice (though often without careful explanation), courts have found a protective duty in somewhat broader and varied circumstances, employing mixed rationales of patients' dangerousness and physicians' ability to take protective steps; hence the choice of the word "hazard" in the above typology, rather than "special relationship."

The cases include physician relationships with non-hospitalized patients with contagious diseases (in which courts have imposed a duty to warn caregivers and family members of their risk) (*see* Bradshaw v. Daniels, 854 S.W.2d 865, 871 (Tenn. 1993) (reviewing cases)); dangerous psychiatric outpatients (beginning with the famous case of *Tarasoff, supra,* which, though novel because it applied the duty to protect third parties to a psychotherapist with a dangerous outpatient, actually relied on the contagious disease outpatient cases just mentioned as precedents); and patients with medical conditions that impair their driving ability. Some modern courts suggest that *any* physician-patient relationship is "special" and triggers a protective duty, without reference to the type or degree of hazard. *Tarasoff, supra*; *Bradshaw, supra*. This formulation imposes an open-ended

physician duty to protect third parties against hazards of any kind. Also, where the physician actually *creates* the hazard, as for example in providing medication that temporarily impairs driving ability, courts sometimes eschew the need to find a "special relationship."

A few courts have moved beyond the requirement that the patient present some personal physical hazard to a third party, holding instead that the physician's duty is triggered merely by learning or suspecting that someone else is in danger, from whatever source. Thus in Bradshaw v. Daniels, 854 S.W.2d 865 (Tenn. 1993), where a man died from a *non*-contagious disease (Rocky Mountain Spotted Fever), the court nonetheless held his physician had a duty to warn the man's wife, who subsequently also died from the disease, that she too was at risk based on the fact that the ticks that transmit the disease to humans tend to "cluster." Two cases, harbingers of an emerging issue, have found that physicians have a similar protective duty running to family members of individuals with heritable genetic conditions. Safer v. Estate of Pack, 677 A.2d 1188 (N.J. Super.A.D. 1996) (physician may have duty to warn family members); Pate v. Threlkel, 661 So.2d 278 (Fla. 1995) (duty to family members recognized; dischargeable by advising patient). This problem is especially complex given the potential tension between such a protective duty and a "right *not* to know" that many persons may assert in order to avoid the emotional implications of confronting their susceptibility to a family history of genetic problems. To resolve the tension perhaps physicians will need to develop

generic questions for family members about their wish to acquire genetic knowledge based on family history, in advance of any actual disclosures—though even doing that much may well breach the patient's confidence or a right not to know.

The continuing extension of a duty to protect third parties may be animated by the "rescue" ethos discussed in sec. A.2.b, *supra*: in some situations a third party's protection simply exerts a

sufficiently strong moral claim that courts will recognize a duty to act, notwithstanding traditional tort rules and the strong competing value of patient confidentiality.

(b) Foreseeable Plaintiffs

The health care provider's duty to accurately detect, or diagnose, the hazardous condition is measured by a professional negligence standard—whether the provider "knew or should have known" of the condition and its dangerousness—an inquiry that normally requires expert testimony. In the *Tarasoff* context (psychiatric violence), many psychiatrists and psychotherapists doubt the profession's, and their own, ability to predict dangerousness accurately. The standard, however, does not require an *accurate* prediction, but only one arrived at *non-negligently*—i.e., through adherence to professional standards of care.

The physician's duty, once recognized, runs to reasonably foreseeable plaintiffs. In many cases this will be only one, or a few, persons of known identity. For example, with psychiatric dangerousness there is frequently a threat to a specific individual (*Tarasoff*).

With contagious disease, family members or caregivers may be the primary persons at risk. In such cases, the imminence of the harm and the reasonableness of imposing the protective duty are perhaps the clearest and the most readily understood.

However, the individuals in the class of prospective plaintiffs need *not* be actually known, or even personally identifiable, in all cases. An impaired driver, for example, threatens all other drivers, passengers and pedestrians on the road. It is *not* the case, however,

that a physician's duty to protect third parties runs to the world at large; there will generally be some limit of reasonableness under the circumstances that identifies the extent of the protected class. Indeed, this variation in *who* must reasonably be protected is related to the final element: *how* protection is to be accomplished.

(c) Discharge of the Duty to Protect

Recall that steps short of breaching patient confidentiality may often suffice to meet the physician's protective duty. For example, an elderly patient with badly impaired vision may be privately persuaded by the physician to relinquish his car keys; an HIV-positive patient may be instructed regarding safe sex or abstinence and the avoidance of needle sharing; a man with a genetic condition might be told of its familial nature and left to his own judgment with respect to family disclosure. On the other hand there are certainly cases where the protective duty might *only* be met by a breach of confidentiality: a

vision-compromised patient who will not voluntarily stop driving; an HIV-positive patient who refuses to adopt safe sexual practices; or a psychiatric patient who intends to harm a former lover. Here the duty of confidentiality and the duty to protect come unavoidably into conflict, generating a *Scylla* and *Charybdis* of competing legal obligations.

Even here, the precise *form* in which confidentiality is broken may well vary with the circumstances, and "warnings" to specific individuals may be neither necessary nor useful. In the case of the vision-impaired driver, the physician might talk to the spouse or children or call the motor vehicle division, but it would be impossible (and thus vain) to "warn" all other drivers and

pedestrians. In the case of the HIV patient, the physician might contact the HIV patient's lover or call the state health department (which, in some states, then assumes responsibility for "partner notification" or other contact tracing). Indeed it is especially logical for state public health authorities to take over the duty to protect in the context of transmissible diseases like HIV, since there may be multiple partners or exposed individuals, whose identity the physician could not reasonably be charged with knowing or ascertaining. In the case of the psychiatric patient, the physician might directly warn the potential victim or contact the police. As *Tarasoff* recognized, discharge of the protective duty—even where it necessarily involves the breach of confidentiality—does not require a warning *per se,* but instead requires "whatever . . . steps are reasonably necessary under the circumstances."

This, of course, is not entirely reassuring, given the ambiguity of what constitutes the "correct" choice and the potential liability for either decision. This is an area where it would be helpful for the law to confer qualified immunity to providers who, in good faith, follow *either* course. Thus a physician would be protected against both the patient's claim of breach of confidentiality and the third party's claim of failure to protect by a "buffer" rule that required him only to demonstrate (for example) careful consideration, investigation of options, and perhaps an (anonymous) ethics consultation, with respect to the most advisable course of conduct. Some state legislatures have enacted just such a rule in the specific context of HIV-related behaviors. The rule might be extended legislatively to other subjects. Alternatively, in common law litigation, a court might judicially create such an immunity—or, if it believed one outcome superior to the other, could nonetheless

announce a liability rule *prospectively* so as to avoid unfairness in the initial case.

3. INFORMED CONSENT

Probably no medico-legal doctrine has received more scholarly attention in the past several decades than the law of informed consent. This is due in part to its central role in defining the legal content of the treatment relationship.

The core value underlying the law of informed consent is autonomy. By requiring a physician to disclose information meant to enable the patient to choose knowledgeably among reasonable medical

alternatives, informed consent seeks to place patients in control of the course of their medical treatment. The rules reflect the agency principles that underlie fiduciary law: the physician is agent who supplies information and advice to patients, enabling them to decide what treatment to order according to their personal preferences and interests.

It is commonly observed that there is considerable tension between the doctrine and aspirations of informed consent law as articulated and advocated by judges, lawyers and legal scholars committed to autonomy, and the real world of medical practice which recognizes competing values (such as beneficence) and, in many cases, doubts the efficacy of informed consent law to achieve its goals. Core assumptions can be questioned as an empirical matter, including patients' capacity to absorb the relevant information, whether people actually desire to make their own medical decisions, and the degree to which autonomy is cherished

as the primary value in medical decisionmaking (among different groups and cultures, as well as by particular individuals). Informed consent claims appear to play a fairly small role in overall litigation against physicians for medically-induced harms, and are rarely brought independently of an accompanying malpractice claim alleging that the care actually delivered was substandard. Nonetheless the doctrine is important, both for its formal contribution to the law "on the books" and because it is one of the tools through which the content of the treatment relationship is negotiated.

125

a. Classic Doctrine

Informed consent derives from the intentional tort of battery, but modern informed consent law in almost all jurisdictions is fundamentally a negligence-based doctrine. To recover, a patient must typically show that: (1) the course of treatment followed carried with it an undisclosed risk; (2) the physician's nondisclosure of that risk breached the applicable standard of care owed to the patient; and (3) the undisclosed risk caused the patient's injury, in both a physical sense (by materializing) and a behavioral sense (in that, with proper disclosure, the patient would have made a different treatment choice, thus avoiding the harm). The following discussion briefly surveys these elements. The other essential preconditions for a valid informed consent are that the patient must have the requisite mental capacity, and consent must be voluntary. Note at the outset that the presence or absence of medical negligence in the *performance* of treatment is irrelevant to the claim; informed consent rests on a separate theory, and separate proof.

(1) Risk: What Information Must Be Shared?

Generally speaking, physicians must disclose the patient's diagnosis; the nature, purpose, and probability of success of the proposed treatment; the salient risks accompanying the treatment, including risks arising from the patient's particular medical susceptibilities; and the alternatives, including *their* risks, consequences, and probability of success. One hopes that these are precisely the issues that all

physicians consider when deciding the best course of treatment, so requiring physicians to convey this to their patients follows directly from the doctrine's theoretical goal of giving patients the benefit of their physicians' superior knowledge and expertise. In reality, however, physicians, like others who practice in a complex profession, do not deliberate explicitly on each factor that might influence each decision. Instead, they adopt "clinical heuristics" or rules of thumb which embed many implicit assumptions, so that informed consent doctrine might alter how physicians think as well as what they say. Often, though, it does not do either, since these disclosures are usually made in writing on forms drafted or approved by lawyers and long since forgotten by the physician.

(2) Negligence: Measuring the Physician's Conduct

Two main rules have developed for assessing the adequacy of the physician's disclosure. About half of states apply a "professional" standard: the physician must disclose information that would be shared by a reasonably competent physician in comparable circumstances. This approach, of course, is quite analogous to the legal requirement for medical or surgical *performance* in malpractice cases, and has the same central characteristic: the law,

in searching for a standard, defers to (normative) medical practice. In informed consent litigation this approach is disadvantageous to patients in two senses. First, as with malpractice, it generally requires expert testimony to prove both what the applicable

professional standard of disclosure is, and to establish that it was breached. Second, there is no relief if the professional standard does not happen to call for disclosure; patient informational expectations, however reasonable, are theoretically irrelevant.

Some commentators have argued, and some courts have held, that the medical profession itself has adopted sufficiently patient-oriented disclosure norms that disclosures will routinely be adequate even under the professional standard. Others defend the "professional" rule on the view that it protects a physician's judgment about how to properly allocate time between informational disclosure and medical or surgical treatment (though that tension is not an inevitable one, since nonphysician providers can also disclose and counsel). In any case there is this irony in the professional standard: a remedial doctrine, founded on the perceived need to enhance autonomy, defers to the collective judgment of those whose behavior generated the need for reform.

A similar number of jurisdictions adopt a "patient-oriented" disclosure standard, which measures the adequacy of physician disclosure not by normative physician practices but by patient needs: all information "material" to the decision of a "reasonable patient" must be "unmasked." Canterbury v. Spence, 464 F.2d 772 (D.C. Cir. 1972). The core of this approach is the notion that, once armed with the requisite information, the decision about what

medical course to follow is a personal, value-laden, nontechnical one, rather than one

dependent upon medical expertise. Many commentators believe this rule better reflects the autonomy-based essence of the doctrine.

This lay-oriented standard avoids the need for expert testimony on the adequacy of disclosure. However, expert testimony may still be needed on causation and on whether the undisclosed information constitutes a risk of, or an alternative to, the treatment. States following this approach generally apply an objective ("reasonable patient") standard of materiality, rather than a subjective ("this patient") one; this is designed to avoid holding physicians liable for failing to accurately guess the unknowable informational needs of particular patients. (Courts say this is consistent with the "foresight" rather than "hindsight" requirements of "orthodox negligence doctrine." *Canterbury*. Query, though, whether it makes sense to measure—as this rule does—the adequacy of A's conduct by the *expectations* therefor of a "reasonable B.") Moreover, it is not obvious how providers are to decide, in advance, that a particular risk would be "material" to even a "reasonable" patient; by its nature, this standard leaves many questions for the jury, which entails great uncertainty for providers.

(3) Causation

There are two dimensions of causation in an informed consent case. First, the patient must actually be harmed by the undisclosed risk. Proof of this often requires expert testimony. Second, the plaintiff must show that the risk's disclosure would

have led to a different medical decision, thereby avoiding the harm. This non-technical question does not require expert testimony. Most courts ask whether a "reasonable" (rather than *this*) patient would have made a different medical choice, in order to avoid the hazards of self-serving, *post hoc* patient testimony; in this approach the patient's own testimony may be relevant but not dispositive. One can argue, though, that the usual vehicles for truth-ascertainment (cross-examination, the existence of prior inconsistent statements, testimony of knowledgeable others and the like) would adequately counter this risk, and that the choice of this objective-patient standard takes away much of the autonomy that the doctrine otherwise promises.

(4) Exceptions to the Duty to Disclose

In several circumstances courts have recognized (often in dictum) exceptions which, in essence, privilege nondisclosure. It is generally the physician's burden to prove such a privilege or defense. By and large these exceptions are narrowly construed because of their capacity to undercut the doctrine. They include, first, undisclosed risks that are "common knowledge," and second, risks of which the particular patient is already aware. There is also no duty to disclose in emergencies, particularly where the patient is incapacitated, though even here "proxy" consent by relatives is advisable. Some courts also recognize a "therapeutic" privilege, when disclosure of the usual information would be so damaging or upsetting emotionally that it would "menace" the patient's well-being. *Canterbury*,

however, cautions against broad recognition of such a principle lest it swallow the rule, and against the medical paternalism of invoking

it as a means of imposing "needed" treatment on the patient that the physician fears would be rejected if fully described. Finally, a few courts have observed in passing that autonomy, taken seriously, allows waiving the right to informed consent, since only an ironically paternalistic view of informed consent would *force* the doctrine's customary vision of autonomy on an unwilling patient.

b. Conflicts of Interest and Fiduciary Principles

Though courts very often *characterize* the treatment relationship as a fiduciary one, their actual reliance on fiduciary doctrines to impose liability has been somewhat selective. However, in a famous case the California Supreme Court held that a physician must disclose economic or research interests "unrelated to the patient's health" that might affect the physician's judgment. Moore v. Regents of the University of California, 793 P.2d 479 (Cal. 1990). The court noted that this result is supported both by informed consent law, and more directly by the conflict of interest prohibition imposed by fiduciary law. In *Moore*, the physician had an interest in developing biotechnology products from cells taken from the patient's diseased spleen. The court held this should have been disclosed prior to surgery and during post-surgical care, especially since some of that care had no direct therapeutic purpose.

Fiduciary doctrine also influenced a ruling under ERISA (which incorporates trust law) that a managed care company must disclose to its members the financial incentives it uses to influence primary care physicians' decisions about when to refer to specialists. Shea v. Esensten, 107 F.3d 625 (8th Cir. 1997). Fiduciary or informed consent doctrine might also require these financial disclosures from

physicians, although this has never been done under traditional fee-for-service arrangements, which create incentives to over-treat. However, those incentives are more obvious. Fiduciary law sometimes considers a conflict of interest to be so debilitating that it is prohibited regardless of disclosure and consent. However, the Supreme Court held that ERISA's fiduciary principles do not prohibit physicians from owning an HMO, through which they might profit from withholding expensive treatment to their patients. Pegram v. Herdrich, 530 U.S. 211 (2000). The Court noted that state law addresses such matters, and so declined to create a new body of federal common law that would determine which forms of physician incentives are impermissible. State law, for its part, has also taken a hands-off approach, so far. Neade v. Portes, 739 N.E.2d 496 (Ill. 2000), for example, cited *Pegram* in support of its refusal to recognize a cause of action for breach of fiduciary duty based on a patient's claim he was harmed by a physician's failure to refer him to a cardiologist, influenced in part by a secret medical incentive fund that rewarded cost-savings. The court reasoned that such matters should be resolved simply through a medical negligence action. This

leaves mainly to regulatory law, discussed in Chapter 1.D.3.b, which financial arrangements in medicine are banned or permitted.

c. Emerging Applications of Informed Consent

Traditional informed consent cases involve medical or surgical risks. Courts have also been asked to include in the disclosure obligation novel kinds of risks, such as the risks of doing nothing, the economic factors just mentioned, or particular traits of the individual physician. These cases generally arise in "patient-

oriented" jurisdictions since its materiality standard allows much more freedom to innovate than does professional custom.

Truman v. Thomas, 611 P.2d 902 (Cal. 1980) is the leading ruling that informed consent applies to a patient's refusal of a diagnostic test. The court sent to the jury a case where the physician failed to impress upon his patient forcefully enough the potential consequences of foregoing the Pap smear he offered to her. In this "informed refusal" situation, the risk arises from the *lack* of a medical procedure rather than its performance. Physicians complain that this rule sets no limits on how aggressive they must be in convincing reluctant patients.

Another developing area is the disclosure of a physician's individual skills. In one case a surgeon performed a complex and risky surgery to clip a patient's brain aneurysm, which went poorly resulting in quadriplegia. The patient did not pursue either surgical negligence nor nondisclosure of the risks of the procedure, but instead argued that the

physician wrongfully withheld information (and may have actually provided misleading answers to the patient's questions) about his lack of personal experience performing this procedure. Evidence indicated, but the patient claimed not to know, that less experienced physicians have higher surgical morbidity and mortality and that a more sophisticated and resource-intensive hospital was available within 90 miles as an alternative site. The court ruled these considerations are material to a reasonable patient's decision and upheld the trial court's admission of this evidence under that theory. Johnson v. Kokemoor, 545 N.W.2d 495 (Wis. 1996). However, there was a suggestion in the case that the physician

misrepresented his credentials, and other courts have ruled there is no basis for suit for mere nondisclosure of experience or success rate, absent misrepresentation. In support of this compromise, physicians note that the stricter position would keep less-senior practitioners from gaining necessary experience and might stigmatize even senior practitioners who attract the most difficult cases, which are more prone to complications. This issue will become increasingly controversial as more provider-specific outcomes measures become available.

A related and sensitive topic is disclosure of personal, non-technical physician characteristics. An HIV-infected physician offers a riveting test case. The statistical risk of a doctor transmitting HIV is extremely low, but public anxiety over the possibility is understandably high, and the consequences of infection are devastating. One decision struck the

balance in favor of the patient, requiring an HIV-positive surgeon to disclose his status. Behringer v. Medical Center at Princeton, 592 A.2d 1251 (N.J. Super. Ct. 1991). Similar questions arise for physicians with substance abuse problems. Physicians argue that these issues are better resolved through hospital privileging decisions, medical licensure, professional ethics, or the threat of malpractice litigation.

In a significant decision pointing the opposite direction, the California Supreme Court affirmed a jury finding that physicians are not bound to disclose a cancer patient's (short) life-expectancy in order to enable him to put his financial and business affairs in order. The court held that, despite the physician's fiduciary status, informed consent law does not require disclosure of risks to

nonmedical interests nor of everything a patient might literally want to know. Arato v. Avedon, 858 P.2d 598 (Cal. 1993). Considering that California is historically among the most liberal jurisdictions in developing and applying the doctrine, this case is an indication of the law's ambivalence about the future scope and application of informed consent law.

4. MODIFYING THE TERMS OF THE TREATMENT RELATIONSHIP

Recall that the formation of the treatment relationship is essentially contractual in nature. Once established, however, tort and fiduciary law generally govern the parties' conduct and their mutual obligations. This section explores the extent

to which patients and physicians are free to modify those rules and define the terms of their relationship in accordance with their own preferences.

Courts have generally refused to enforce agreements with patients by which health care providers try to waive their liability for negligence. The leading case is Tunkl v. Regents of the University of California, 383 P.2d 441 (Cal. 1963), in which the California Supreme Court concluded that a hospital's exculpatory agreement with a patient, signed at admission, bore all the indicia of an unconscionable adhesion contract.

Tunkl and similar cases do not, by their terms, preclude liability waivers that fall short of full exculpation, and in fact courts are likely to enforce releases from liability where the care provided departs from standard medical practice for good reason, as when a patient leaves the hospital early against medical advice, or refuses

recommended medical treatment. In this connection, consider Shorter v. Drury, 695 P.2d 116 (Wash. 1985), holding that a document signed by a Jehovah's Witness surgical patient, releasing providers from responsibility for consequences "due to my refusal to permit the use of blood," constituted an enforceable assumption of the risk of an otherwise-avoidable death, rather than an unenforceable exculpation of negligence.

Courts have also responded favorably to agreements changing the forum or mechanism of dispute resolution. For example, HMOs may require their members to arbitrate rather than litigate

medical negligence claims, at least where this is agreed to through bargaining by a powerful representative (a large employer) and there is a choice of alternative plans. Madden v. Kaiser Foundation Hospitals, 552 P.2d 1178 (Cal. 1976). On the other hand, "point of treatment" arbitration agreements, presented for signature upon hospital admission or at the doctor's office, are less likely to receive judicial approval because of concerns about their fairness, though these decisions too are not uniform.

Agreements to *alter* the prevailing standard of care, rather than to waive it entirely as in *Tunkl*, are more difficult. If notice is adequate and there is some choice, should HMOs (for example) be allowed to contractually bind enrollees to accept a lower-than-normal standard of care (e.g., anything above "gross negligence") by plan providers, as a cost-containment mechanism that would benefit enrollees by reducing premiums? The law on such questions is not well-developed, but these questions could become a focal point of controversy if medical standards became subsumed within managed care contractual arrangements.

[1] A malpractice-type standard, in fact, is not inconsistent with EMTALA's statutory language for "appropriate" screening "within the capability" of the institution. The former reflects malpractice law's normative, objective content; the latter reflects its recognition that the standard of care may vary with a hospital's treatment category and, perhaps, available resources.

PART II

THE LEGAL STRUCTURE OF HEALTH CARE DELIVERY

The pressures of the cost, coverage and reform described in Chapter 1 are pushing the health care industry to develop organizational forms and relationships that do not fit easily into preexisting legal categories. As a consequence, health care corporate and regulatory law remains a flourishing practice area that requires sophistication in subjects as diverse as antitrust, tax, licensure, and insurance regulation. The materials in the second part of this book examine these and other private and public law doctrine that have the greatest relevance to the unique legal problems presented by the structure and functioning of the contemporary health care delivery system.

CHAPTER 3

HOSPITAL STRUCTURE AND REGULATION

The prototypical health care institution is the general, acute care, medical-surgical hospital. Hospitals come in many shapes and sizes. Most are private and non-profit, although many are run by government entities or owned by investors. Some are quite small, with 50 beds or less; others are 1000-bed behemoths. Community hospitals offer more basic services such as routine child birth and simple operations while major medical centers and teaching hospitals strive to have the most comprehensive and state-of-the-art programs and technology available. Other hospitals specialize in only a limited range of medicine, such as mental health or cancer. Nevertheless, hospitals of all types share important common features. There are also common features among all medical institutions. Although we seldom mention nursing homes, home health agencies, diagnostic clinics, and ambulatory surgery facilities in this book, often times when we refer to hospitals we could equally well include these other medical facilities. Finally, the border between health care insurance and health care delivery is less distinct than it once was due to innovative arrangements like HMOs, which both deliver and finance medical care. Therefore, many of the laws that affect traditional facilities also apply to HMOs or to joint ventures between doctors and hospitals.

This chapter explores the basic legal environment in which traditional hospitals and other health care facilities are organized and operated. It addresses the bodies of law that are most familiar to health care lawyers, and for the most part it regards these facilities in their simplest structure. Although this chapter considers the obvious extensions of this basic legal doctrine to HMOs and other more complex ventures, most of the cutting-edge legal developments that respond to these innovative structures are taken up in the subsequent chapters.

A. HOSPITAL AND FACILITY REGULATION

1. LICENSURE, CERTIFICATION, AND PRIVATE ACCREDITATION

Hospitals and other health care facilities such as nursing homes are so heavily regulated that they are sometimes thought to approach the status of public utilities. One hospital once toted up that it must submit reports to and comply with rules set by over three dozen government authorities and a half dozen private bodies. The primary authorities are: 1) state licensure, 2) private accreditation, and 3) certification for participation in government insurance. These three sources are each legally distinct, but their substance and processes are intertwined to a considerable extent.

Since the mid-1900s, virtually every state has regulated the operation of hospitals, nursing homes, and similar facilities through licensure statutes and regulations. Hospital licensure provisions typically

read like a gigantic building code for the hospital industry, specifying a host of architectural, safety, and sanitation minutia as a

condition for issuing or renewing an operating permit.

Private accreditation of hospitals and other health care facilities overlaps to a significant extent the function performed by state licensure. The Joint Commission on Accreditation of Healthcare Organizations, which is referred to as the "JCAHO" or the "Joint Commission," is a private accreditation body maintained jointly by the AHA, the AMA, and two other physician groups. Its accreditation standards impose detailed organizational and procedural standards for the structure and operation of each hospital department. The JCAHO wields enormous authority and influence because virtually no hospital of respectable size risks the business consequences of jeopardizing its accreditation status. Many states effectively delegate their licensing function to the JCAHO by incorporating its standards by reference.

Similarly, hospitals and other facilities must be certified as fit to participate in Medicare and Medicaid, under standards very similar to those for licensure and accreditation. Accordingly, the federal Medicare program automatically deems all JCAHO-accredited hospitals as meeting certification standards for participation. One can question whether it is an appropriate public policy to abdicate regulatory oversight to the industry itself, but "deeming" has been upheld as constitutional. Cospito v. Heckler, 742 F.2d 72 (3d Cir.1984). Certification

standards are somewhat more demanding for nursing homes under Medicaid because this has been seen as an area were traditional state oversight mechanisms were lax. In response to litigation, *see* Estate of Smith v. Heckler, 747 F.2d 583 (10th Cir. 1984), HHS has imposed an extensive set of regulations that govern in considerable

detail the treatment plans, living environment, legal rights, and human dignity of nursing home patients. They are enforced primarily through state licensure officials.

This overlapping regulatory structure raises a host of legal issues, only a few of which will be touched on here. (Among the omitted issues are constitutional and procedural challenges to adverse decisions.) The first issue is jurisdiction. Facility licensure typically excludes physician offices under the premise these fall under the jurisdiction of physician licensure. But, when is a physician's office really a facility? Consider, for instance, freestanding urgent care centers, colloquially known as "Doc-in-Boxes." These store-front medical clinics cater to no-wait, no-appointment medical needs of an urgent nature, short of life-or-limb threatening conditions. They provide a convenient and less expensive alternative to hospital emergency rooms for conditions such as broken legs, bad cuts, and sudden illnesses. It can be argued that urgent care centers are nothing more than glorified doctors' offices, which traditionally have not been covered by facility licensing laws. However, states have amended their statutes to cover these and other novel delivery arrangements such as ambulatory surgery clinics.

A second set of issues concerns how these regulatory authorities attempt to define and oversee the quality of care. Public policy theorists distinguish among three different ways to measure quality of care: *structural measures* such as corporate and managerial organization and the composition of relevant committees; *process measures* such as protocols and forms that must be followed by professional staff to avoid mistakes and catch errors; and *outcomes measures*, which look at how patients are actually doing. Critics charge that licensing, accreditation, and certification are directed

almost exclusively to structural and process measures, which produces an excessive focus on red tape and busy work for administrators.

Attempts to invoke bottom-line outcomes measures are underway, such as infection rates, death rates, and patient satisfaction, but they encounter persistent difficulties over the methods of measurement and comparison, so that low-scoring facilities are not unfairly penalized by bad luck or because they happen to attract more difficult or serious cases. Both the Joint Commission and the Medicare/Medicaid certification standards now require that hospitals and other institutions adopt outcomes measures of quality, as does the National Committee on Quality Assurance (NCQA), which accredits HMOs. Ultimately, however, this push for outcomes measures may devolve into another set of structural or process measures. Licensing and accrediting organizations find it difficult to impose any absolute performance standards, since numerous factors, many out of the facility's control, affect how

patients actually do. Therefore, most of these new standards require simply that facilities take and monitor these performance measures, not that they achieve any particular score.

2. CERTIFICATE OF NEED LAWS

A more specialized regulatory regime that exists in most states requires hospitals and other health care institutions to obtain a certificate of need ("CON") from a government agency before constructing new facilities, purchasing major medical equipment, or instituting new health services. CON laws arose from the National Health Care Planning and Development Act of 1974, which required states to adopt CON regulation in order to receive

certain federal health care funding. In 1987, though, Congress became disenchanted with the CON regulatory approach (for reasons explained below) and repealed the federal mandate, leaving states free to depart from the federal model. Consequently, a number of states have scrapped their CON laws entirely and a number of others have substantially loosened their regulatory reigns. Still, CON has a significant presence in the majority of states and is expected to remain as a permanent fixture in the health care regulatory apparatus for some time to come.

Certificate of need laws are a response to the excess capacity in capital resources. Hospitals commonly have far more space than they require, frequently operating at 50 percent capacity or lower. The hospital industry is also notorious for its redundancy of technology: if one hospital acquires

the latest technological gadget—formerly, a CAT scanner, later a magnetic resonance imager (MRI), and now a positron emission tomographer (PET)—then immediately all of the other hospitals in town set about making duplicative purchases. The same is true for glamorous services such as heart transplant programs.

Ordinarily, market forces penalize such overinvestment of capital resources because a firm with excess capacity must charge higher prices to service its debt or to provide a competitive return on investment equity. More conservative firms are then capable of quickly undercutting the high spender, resulting in a net decrease in income. In health care, however, higher prices do not automatically cause a loss of business under traditional, open-ended, cost- or charge-based reimbursement.

A second inflationary factor in capital spending is the nonprofit status of much of the hospital industry. The only use that charitable hospitals are permitted to make of their earnings is to plow them back into the facility. The prestige of non-profit managers is measured not by the profitability of the enterprise but by its size and presence in the community. This leads to what is sometimes called the hospital industry's "edifice complex"—an almost obsessive desire to build and spend.

If all we faced were a single layer of fat atop an otherwise lean hospital industry, our problems would not be so severe, but an important phenomenon amplifies the effect of this profligate capital spending. Health care appears to be ruled by what is

known as "Roemer's law," which states that empty beds generate increased demand for services. Roemer was the first to identify and explain the apparent paradox that health care utilization tends to be the highest where there is the most unused capacity. When hospitals expand, they urge their medical staffs to think of more inventive ways to use the facilities, and once the facilities are fully used, hospitals expand once again. The fundamental point of Roemer's law is that more than the costs of bricks and mortar are at stake: capital expenditures drive up operating costs, which lead to increased reimbursement and then further expansion, in a never-ending spiral.

CON laws are designed to curb these excesses by requiring hospitals and other health care facilities (such as nursing homes, ambulatory surgery clinics, and home health agencies) to demonstrate a need for new projects that involve a substantial expenditure. The exact expenditure thresholds and project

descriptions vary widely from state to state, but generally speaking the amounts involved must be $1 million or higher.

Hospitals have found various inventive means for circumventing the reach of CON regulation. The most prominent technique is to arrange a new project under the auspices of a physician rather than a hospital because most CON laws apply only to health care institutions and specifically exclude expenditures by physicians in private practice. In one early case, a group of doctors was successful in purchasing without approval a magnetic resonance

imager, a multimillion dollar piece of sophisticated diagnostic equipment usually found only in major medical centers. Boulware v. State, Dept. of Human Resources, 737 P.2d 502 (Nev.1987). However, some states have amended their statutes or regulations to encompass projects of this nature that are intended to serve hospital patients.

When a project—say, the construction of a new hospital—is subject to review, the controversy shifts to how the need for additional hospital capacity (measured by number of beds) is properly determined. Bed-need methodology is usually spelled out, either in a State Health Plan or in CON regulations, in a manner that establishes a mathematical formula for comparing the existing stock of beds with the projected demand. For instance, a simple formula might call for a maximum of 4 beds per 1000 population in a designated geographical region. A more refined formula might set the ratio according to the historical utilization experience of each local population group to account for differences in age or health, or might use more sophisticated techniques to account for the net migration of patients in and out of the geographical area.

If these mathematical formulae fail to establish a net shortage of hospital beds, CON applicants frequently attack the formulae as being unreasonably rigid and tending to ignore other, less quantitative statutory factors such as the quality of services proposed and the increased accessibility to underserved groups. While such attacks are difficult

to maintain, they have enjoyed surprising success in some jurisdictions.

A CON applicant's task is far from over when it identifies an underserved pocket of population. Typically, such an applicant will find itself fending off several other competing applications. These comparative hearings can quickly degenerate into disputes over trivia such as who has the superior parking lot design or who is cutting down the fewest trees. Procedural complexities also abound in the CON review process, so that the process is often lengthy and expensive.

For these and other reasons, several studies have demonstrated that CON regulation has virtually no effect on health care investment or expenditures. This failure is due primarily to several shortcomings in the design and implementation of CON laws. Even as designed, CON laws are very limited in scope. First, they address only the major capital *costs* of health care; hospitals are still free to *charge* whatever they want, and they are free to release their spending pressures in other directions such as salaries and other operating costs. Second, "need" for new facilities is usually measured in terms of current—*i.e.,* inflated—treatment patterns. Therefore, CON laws at best remove only the outer layer of fat, that is, the existing excess capacity; they are inherently incapable

of reversing the built-in inflationary base of treatment intensity caused by the "Roemer cycles" of past decades. Moreover, CON laws are constitutionally incapable of eliminating increases in service intensity. With "need" as the

149

primary focus, it is impossible to turn down any new facility or technology that provides a conceivable benefit.

CON is more than simply a failed attempt, though; it may have produced net harms. Most obvious are the costs of administering a complex and broad-based regulatory program. Less obvious, but more troubling, are the anticompetitive effects of limiting market entry by new hospitals. Prohibiting new construction absent a showing of need protects established hospitals from competition by new firms that desire to enter overbuilt markets, the very markets where competition is the most needed. Moreover, where new construction is warranted, it has sometimes been observed that CON regulators play favorite to local hospitals over newcomers in choosing among competing applicants. This is not entirely the result of dirty pool; the more established applicant is inherently able to expand at the lowest cost. It is for this reason that certificates of need serve almost like franchises in perpetuity: once a hospital has a foothold in the market, it is in the naturally favored position to continue to expand to meet future growth in need. This protectionist character of CON laws explains why the hospital industry strongly supports this form of regulatory control.

150

B. HOSPITAL AND HMO MEDICAL STAFF ISSUES

1. MEDICAL STAFF STRUCTURE AND STAFF SELECTION PROCESS

The hospital medical staff is an institution unique to North America. Elsewhere in the world, hospitals employ a select group of specialist physicians who practice exclusively in a single hospital under a salaried arrangement. In the United States (and Canada), the tradition has developed that virtually all physicians practice in more than one hospital and they do so independently of the hospital. Doctors are neither paid by hospitals nor do they pay for the privilege of using hospital facilities. The relationship between doctor and hospital is symbiotic, each sustaining the other through an implicit exchange of benefits—the hospital provides a doctor's workshop and the doctor provides a supply of patients.

This fundamental division between hospital and doctor pervades the health care delivery system. A hospital patient (or her insurance company) will receive a bill for physician services that is separate from the bill for hospital services. The core of Medicare is split into two parts, Part A for hospital (and other institutional) costs and Part B for physician charges. Blue Cross/Blue Shield plans, which are the largest component of the health insurance industry, reflect the same dichotomy.

The only formal manner in which hospitals and doctors interact is through the hospital medical staff. The medical staff is the collection of doctors that

enjoys admitting privileges at a hospital. The significance of "admitting privileges" is understood only by realizing that, in a formal sense, hospitals do not admit patients directly; instead, they receive patients that are admitted by practitioners on the medical

staff. Thus, although American hospital medical staffs are much more "open" than their European counterparts, it is not the case that literally every doctor may admit patients to every hospital. Only those practitioners who meet the criteria established in the hospital's medical staff bylaws are allowed to join.

The medical staff bylaws are an organizational document that is separate from the hospital bylaws. JCAHO accreditation standards emphasize that the staff bylaws may not be amended unilaterally by the hospital (or by the medical staff). The medical staff's effective veto power over selection criteria further reinforces physicians' control over the credentialing process that is explained in the following section. These organizational patterns developed by custom in the early twentieth century and became formally institutionalized in the mid-century through the standards set by the JCAHO.

In addition to admitting privileges, medical staff membership also determines "clinical privileges," that is, which department of the hospital a physician is allowed to join and what procedures within that practice area the physician may perform. Also, there are various categories of staff privileges that reflect the degree of regularity of a physician's practice at

the hospital: active staff, courtesy staff, and visiting staff.

The process of selecting and periodically reevaluating medical staff members is referred to as "credentialing" or "peer review." Because hospital credentials are based on the practitioner's medical competence, the hospital delegates much of the evaluation and decision making authority to the existing medical staff members themselves (hence the name "peer review"). For new applicants,

various medical staff committees first conduct a fact-gathering process and the staff then votes and sends its decision to the hospital board of governors as a recommendation, which the board usually accepts. Existing staff members are reviewed every two years through a more streamlined process. If either review process produces an adverse finding, the initial or renewal applicant is entitled to a formal, evidentiary hearing to contest the finding. If the finding is sustained, the hospital can revoke, suspend, or limit the privileges of existing staff members.

There are three important qualifications to the generalized medical staff model that has just been depicted. First, physicians are not the only medical professionals that potentially can hold staff privileges at a hospital. Beginning in 1984, the JCAHO accreditation standards were revised to allow hospitals to admit psychologists, nurse midwives, chiropractors, podiatrists, and other so-called allied health professionals—licensed medical practitioners whose practice areas are restricted to some limited segment of medicine. The second

qualification relates to certain hospital-based specialists. Hospitals usually find it convenient to employ or contract exclusively with a limited number of radiologists, anesthesiologists, pathologists, and sometimes emergency room physicians. Unlike most practicing doctors, these hospital-based physicians typically do not admit patients and they have direct financial arrangements with the hospital, either as partners, independent contractors, or employees. Finally, teaching and government hospitals depart from the traditional American model by employing large portions of their medical staffs.

Another important development is the replication of the medical staff structure within other medical institutions, particularly, within HMOs. Although HMOs do not give physicians the same autonomy over financial matters or the content of the bylaws, they do replicate the credentialing and peer review processes in deciding which physicians to admit to and retain in the network. Accordingly, many of the issues raised in this chapter with respect to hospitals also exist for HMOs, but in a somewhat altered legal framework.

2. ECONOMIC CREDENTIALING, EXCLUSIVE CONTRACTS, AND INSTITUTIONAL CONTROL

Some hospitals have grown increasingly disenchanted with the traditional medical staff model under which physicians remain financially and organizationally autonomous from the institution. These hospitals have attempted to fundamentally restructure themselves into a more

classically hierarchical structure in which practicing physicians are subservient to top administrators. (Naturally, there are both physician and lay administrators.) For the most part, they have been largely unsuccessful because of the numerous ways in which the independent medical staff is firmly institutionalized both in practice and in law. Legally, medical staff independence is virtually assured because it is written into the existing corporate bylaws. Most courts have ruled that medical staff bylaws constitute a contract that cannot be unilaterally amended by the hospital administration. Even if this were not the case, altering the traditional arrangement might be found to violate JCAHO

accreditation standards, which are embodied in licensure and certification law.

Failing wholesale restructuring, hospitals have attempted to accomplish de facto restructuring by inserting new criteria into the bylaws that define medical staff membership, or by circumventing the medical staff credentialing process through exclusive contracts. These efforts are discussed under the term "economic credentialing." The gist of this term is to call attention to the effort to use economic criteria as well as quality of care criteria to determine physicians' relationship with the hospital. For instance, a hospital might attempt to exclude from the staff or terminate the contract of a physician who consistently loses money on their Medicare or HMO patients, or a physician who takes profitable business from the hospital by opening an outpatient surgery center.

When hospitals take the direct route of attempting to amend the formal bylaws, they usually do not succeed. If they cannot convince the existing staff to vote in favor of these changes, courts have held that the hospital lacks authority to unilaterally amend the bylaws or to make medical staff decisions outside of the authorized credentialing process. However, hospitals have adopted a different technique with more success, using exclusive contracts. They have limited clinical privileges in specified departments to select physicians or groups under contract with the hospital, as is traditionally done with hospital-based physicians. When a physician's contract is terminated or given to someone else, the physician often complains that there has been a de facto termination of medical staff membership. A number of courts have responded, however, that contract termination leaves medical staff privileges unaffected. Since contracting naturally falls within the

purview of hospital administration, hospitals are able through this technique to limit hospital access using any criteria they wish.

How is this legal position sustainable? Courts draw the technical distinction between the honorific or status aspects of possessing medical staff membership versus the practical opportunity to exercise these privileges. Only the former, they reason, is subject to the credentialing process; the latter is under the control of hospital administration. This may appear at first to be an unconvincing hypertechnical distinction, but courts reason it has something of substance. Physicians with staff privileges can still enjoy the reputational benefits of

this status, even if they have limited access to the hospital. And, hospitals must be afforded some opportunity to control physical access lest some clinical areas becomes overcrowded or uncoordinated. Consider, for instance, a hospital that allowed any physician who wanted to work in a hospital emergency room at any time. Sometimes, there would be no one there; other times, the emergency room might become like a crowded bazaar in which physicians vie for each new patient who enters the door. On the other hand, this characterization is clearly exaggerated.

In any event, economic and administrative issues are now influencing medical staff decisions much more than they did in the past. No case to date that considers the merits cleanly sustains a hospital's exclusion decision based on criteria unrelated to, much less opposed to, the quality of patient care. However, this is accepted practice for HMOs.

3. INSURANCE DESELECTION AND MANAGED CARE CONTRACTING

One has to wonder how health care institutional terminology initially takes hold and later evolves. One such puzzle is the term "deselection." This has become the accepted term to refer to decisions by health insurers to drop physicians from their networks. Naturally, this is beginning to produce considerable litigation, as physicians claim they are dropped for no good reason, or for reasons contrary to the public interest such as advocating too

strenuously for their patients or objecting to insurance policies they think are wrong.

At the outset, HMOs avoided the dilemma they saw hospitals in when confronting economic credentialing. Rather than allow medical staff decisions to be made only by a physician-controlled credentialing process, health insurers' managed care contracts create parallel processes which they can choose between. Manage care networks such as HMOs sometimes use a classic credentialing process focused on quality of care, but their physician contracts also provide the explicit right of either party to invoke "no-cause" termination. This allows an insurer to drop a physician without any explanation when it prefers not to air its true reasons. This also allows physicians to leave the network without penalty if they are dissatisfied.

Physicians who are dropped unceremoniously in this fashion allege in lawsuits that no-cause terminations are used to cover up nefarious insurer motives. For instance, HMOs have been accused of dropping physicians whose race or location brings in less desirable patients. In response, courts have split in their rulings. Some courts simply enforce the contractual arrangements as written. Others impose a public policy override, which refuses to

let insurers drop physicians for reasons contrary to the public interest or that would breach the implied covenant of good faith and fair dealing. The leading decision is Harper v. Healthsource New Hampshire, 674 A.2d 962 (N.H. 1996), which sustained a cause of action against an HMO that dropped a physician without

explanation, where the physician alleged the real grounds related to differences of opinion over patient care policies. The law here is influenced by the judicial limitations in recent years to the common law "at will" employment doctrine. This public policy limitation allows physicians who make the proper allegations to obtain a hearing on the actual reasons for termination. They bear the burden of proof, however, and they must show reasons that are more than just arbitrary or mistaken. This is a more demanding showing than that required under the body of law discussed in the next section that governs hospital medical staff disputes.

Before turning to that topic, it is worth considering why "no-cause" termination clauses are so common in managed care physician contracts. One might think that physicians would strenuously object to including these because they shift so much power to the insurer, and so their presence signals that the insurer has disproportionate bargaining power. This is undoubtedly true in some instances, but in others physicians insist on including these clauses. They often enter managed care contracts somewhat wary about how things will turn out and therefore eager to be able to withdraw quickly and easily. Also, if the insurer were to want to terminate them for quality of care reasons, they would much prefer a less explicit process and one that does not result in having to report the action to the National Practitioner Data Bank.

As for the issue of bargaining power, in many markets insurers feel that physicians have the upper,

or at least an equivalent, hand. Most physicians contract with insurers on a nonexclusive basis and so physicians can shift their allegiances similar to the influence they wield over competing hospitals (although they lack the same ability to take their patients with them when they switch insurers). Accordingly, they are often in a position to negotiate aggressively with insurers that recruit them to their networks. This is especially true for popular physicians whose patients will select insurance plans based on their membership in the network.

When negotiating managed care contracts, several other important issues arise in addition to no-cause termination. Primarily, physicians must read these contracts with care to determine what payment and treatment terms they are agreeing to. Often, managed care contracts are written in a manner that gives the insurer carte blanche to bind the physicians to any payment and coverage terms they happen to negotiate with employers or other purchasers. Another controversial issue is liability. These contracts often use indemnification clauses to shift all liability for medical outcomes to the physician, even though some bad outcomes may result from the insurer's own decisions about which treatments are covered.

4. HOSPITAL MEDICAL STAFF DISPUTES

a. Introduction

Long before the insurer deselection controversy, hospitals found themselves in the confluence of

several strong legal and economic currents affecting the medical staff relationship. Under the tort law development known as "hospital corporate liability," a hospital (and potentially its medical staff) is responsible to its patients for the quality of the physicians allowed to practice within the institution. Darling v. Charleston Community Mem. Hosp., 211 N.E.2d 253 (Ill.1965). However, a strong opposing force exists by virtue of the increasing economic importance to physicians of medical staff membership. Access to a hospital is essential to carry on almost any form of medical practice. Physicians who are excluded from the major hospitals (and often there is only one hospital in a community) have their entire professional practice at stake. Moreover, physicians view it as essential to have access to several hospitals in order to compete more effectively for a supply of patients, particularly in recent years with the physician market experiencing what some consider a "glut" of doctors.

Nevertheless, if these were the only interests at stake in medical staff selection, it would be puzzling if courts were to concern themselves greatly with physician complaints about the merits of negative membership decisions. Excluding a qualified physician runs counter to a hospital's fundamental economic interest because hospitals obtain patients mainly through their medical staff members. This economic motive would be expected to serve as an effective check against hospitals screening doctors with excessive zeal. The flaw in this supposition is that staffing decisions are not entirely based on the institution's interests. The existing medical staff,

which has tremendous influence over these decisions, has a precisely contrary interest: it stands to *lose* business when new members are admitted to the staff. This built-in economic bias creates a risk that meritorious physicians will be excluded for reasons unrelated to patient care.

Given the intersection of these three sets of powerful interests—the institution's, the physician applicant's, and the existing medical staff's—it is no surprise that medical staff disputes have boiled over into the courts in great numbers, second only in the medical/legal arena to malpractice litigation. The remainder of this section examines the judicial response to these challenges under the common law. ("Exclusion" refers to either the denial of an initial application or the revocation of existing staff privileges.) Antitrust theories are reserved for Chapter 4.

b. Theories of Judicial Review

A variety of conventional causes of action are available to an excluded physician, but each of these theories is limited in some significant respect. A doctor might allege a violation of due process and equal protection rights under the Constitution, but this theory is limited to *public* hospitals—those owned or operated by state or municipal authorities. At a private hospital, a doctor might contend that his dismissal from the staff violates contractual rights of process and substance contained in the medical staff bylaws. However, this theory is not available to new applicants who have not yet been admitted to the

staff, and some courts have rejected this theory altogether. Under tort law, excluded physicians can allege defamation or tortious interference with contract, but these theories are also laden with

various elements that restrict their application to staff exclusion decisions generally.

Federal statutory law protects against discrimination in the workplace. Courts divide on whether medical staff membership is the functional equivalent of employment. Regardless, this law reaches only discrimination based on race, sex, religion, national origin, or disability. State law is beginning to offer medical practitioners somewhat greater relief through statutory provisions in several states that impose requirements of procedural and substantive fairness on hospitals. Most notably, some states prohibit discrimination based on a practitioner's medical degree, school of practice, or nature of license. These statutory protections remain spotty, however. Consequently, excluded practitioners have sought innovative common law theories that provide more broad-based scrutiny of the procedural and substantive fairness of private hospital peer review.

Excluded physicians have principally attempted to rely on a novel theory that characterizes private hospitals as public facilities. The pathbreaking case is Greisman v. Newcomb Hospital, 192 A.2d 817 (N.J. 1963). There, the hospital maintained a policy requiring all staff members to hold an M.D. degree, which excluded Dr. Greisman, an osteopath, from consideration. The court acknowledged that

Newcomb Hospital was a private facility and therefore not subject to constitutional constraints; nevertheless, it imposed a *common law* duty of fairness in the medical staff selection decision based on its finding that hospitals constitute a "quasi-public" facility. This is a sharp departure from the principle of free association that is

deeply imbedded in our market-based economy, according to which private enterprises ordinarily can decide to do business or not with anyone they like for any reason, absent invidious discrimination or antitrust violation.

The *Greisman* decision attracted much academic attention and a notable following among other state courts. However, several states have refused to adopt the quasi-public characterization of their private hospitals. Conceptually, the problem is in identifying what characteristics should appropriately render a private business "quasi-public" and thus subject to such regulation. The New Jersey court relied on a body of now obscure law. These cases originated from ancient precedents in 15th century English common law that treated common carriers and innkeepers as "businesses affected with the public interest," thereby imposing on them unique public service duties of fairness in their terms of service and prohibiting discrimination against classes of customers. This area of the common law fell dormant about a century ago as it became subsumed by statutory regulation of public utilities. Nevertheless, *Griesman* found that these common carrier precedents still apply in the modern context.

164

These old cases generally applied to common callings that enjoyed monopoly status. For instance, in 15th century England, inns typically were spaced evenly along the roadway at the distance of one day's journey and only one ferry operated at a local river crossing. Sometimes, these common callings also enjoyed support from the sovereign through an exclusive royal charter. These characteristics were viewed as justifying special duties of public service because, if these businesses arbitrarily denied services, the customer would have no alternative source.

Likewise, Newcomb Hospital in the *Greisman* case had effective monopoly status since it was the only facility in the area. And the hospital received ample government support through construction grants, tax exemption, and county funding of indigent care. It might be objected that common carrier law protected only the rights of customers (*i.e.,* patients), not employees (doctors), but the court was careful to observe that, because patients obtain hospital admission only through staff physicians, the discrimination against Dr. Greisman was in effect discrimination against his patients.

However, there are important objections to the quasi-public theory, some of which the courts have articulated and some of which they have not. First, the key characteristic for public facility status in other areas of the law seems to be the *natural* monopoly character of the enterprise. Railroads, telephone companies, and the public utilities have all been regulated as public facilities because of the

belief that they were natural monopolies. While single-hospital communities are certainly common, this is far from the necessary or even predominant industry pattern. Where there are several competing hospitals, or where there is the *possibility* for hospital competition, it is unclear why hospitals require any greater policing than private business decisions in any other important industry.

Second, if public service duties are imposed, this does not extend protection to all physicians, only those attending physicians with primary authority over patient admissions. This limitation would keep the doctrine from applying to hospital-based physicians such as radiologists and pathologists. Third, once the common carrier precedent is invoked, it would seem necessary to follow all of the

ramifications of that precedent, which leads to some surprising and unsettling results. Why stop at physician staff membership; what about other employees? Why stop at admission policies; what about general patient care policies? Observe that ancient English courts scrutinized not only the customer selection policies of common carriers but also the prices they charged. Even more startling, realize that common carriers were held strictly liable for the injuries they caused. These possibilities have not yet been explored in the courts, but they remain pregnant with potential for imaginative plaintiffs' counsel.

c. The Scope of Judicial Review

In states where judicial review is available under the public facility theory, the controversy then shifts to the scope or intensity of review. In *Greisman,* for instance, the court held that a policy excluding all osteopaths is illegal because osteopaths are licensed by the state to practice the full range of medicine and the hospital presented no evidence that osteopaths as a group provide inferior care. The court ruled that it is illegal "to preclude an application for staff membership, not because of any lack of individual merit, but for a reason unrelated to sound hospital standards and not in furtherance of the common good." This statement leaves a number of unanswered questions, though: Are any general, class-based distinctions among practitioners valid, or must anyone with a relevant license be considered on their individual merits? If some general criteria are allowable, what degree of evidence or justification is required to establish the validity of such criteria? And what procedural safeguards are required to insure that criteria are reasonably applied in particular cases?

Greisman encourages hospitals to review physicians on their individual merits. In doing so, what criteria validly count toward merit? The courts, in answer to this question, have articulated a number of legitimate and illegitimate considerations. It is proper to require personal references, to insist on adherence to medical staff rules, to require residence near the hospital, and to require a minimum level of malpractice insurance. It is not proper to require

membership in the local medical society or to require references from current members of the medical staff. The harm of the latter criteria is that they give incumbent physicians the power to "black ball" new arrivals to the community.

When a medical staff takes adverse action against an individual physician based on legitimate considerations, say, lack of professional competence, the court then must determine how closely to scrutinize the evidence supporting that decision. Courts have expressed extreme reluctance to second guess the informed judgment of medical professionals. Representative is this oft-quoted statement:

> No court should substitute its evaluation of such matters for that of the Hospital Board. . . . Human lives are at stake, and the governing board must be given discretion in its selection so that it can have confidence in the competence and moral commitment to its staff. The evaluation of professional proficiency of doctors is best left to the specialized expertise of their peers, subject only to limited judicial surveillance. . . . In short, so long as staff selections are administered with fairness, geared by a rationale compatible with hospital responsibility,

and unencumbered with irrelevant considerations, a court should not interfere.

Sosa v. Board of Managers, 437 F.2d 173 (5th Cir.1971).

168

This deferential attitude does not apply uniformly in all medical staff disputes, however. When there is reason to suspect that an asserted ground for exclusion is in fact a disguise for personal animosity or bias, courts have exercised much more stringent review. A good example is the standard of review for staff discipline based on alleged personality problems. Courts have sustained the exclusion of a doctor with a history of rancor and disruption since poor relations among the staff may impair the quality of service provided the patient. However, medical staff disharmony often grows out of a doctor's legitimate criticism of hospital policy or of other doctors' performance, or it might have its basis in personal or economic matters wholly unrelated to medical policy. Therefore, other courts have balked at the exclusion of physicians simply on the basis of their inability to work with others on the staff; they have required some greater, more specific showing that this inability directly jeopardizes patient care.

Greisman might be read to preclude any rejection of a medical staff applicant other than on an assessment of the practitioner's individual competence. This reading is clearly too broad, however. "Courts should sustain a hospital's standard for granting staff privileges if that standard is rationally related to the delivery of health care." Nanavati v. Burdette Tomlin Memorial Hospital, 526 A.2d 697 (N.J.1987). The problem in *Greisman* is that the hospital made no attempt to justify a rational basis for excluding osteopaths.

Suppose the hospital had argued that M.D.s, as a lot, are marginally better than D.O.s because most osteopaths failed to get into

regular medical school. Would that be sufficient to justify the discrimination? Or would it be sufficient to contend that osteopaths subscribe to a different school of practice and the hospital wished to limit itself to practitioners who all hold the same treatment philosophy?

The answer depends on what degree of scrutiny a court chooses to inject into a review of "rationality." Although constitutional principles do not apply to private hospitals, it still may be helpful in understanding the common law fairness doctrine to recall the analytical framework that courts have developed in their equal protection constitutional jurisprudence. Class-based discrimination against a school of practitioners might be reviewed under a "minimum scrutiny" standard of rationality, which would accept almost any plausible explanation for an exclusionary policy, or it might be scrutinized under a more heightened "intermediate" level of review that requires substantial justification if there is some reason to be suspicious of the medical staff's motives. (A "strict scrutiny" standard that required a "compelling interest" would be inappropriate unless the hospital employed a "suspect classification" such as race, gender or national origin.)

Most courts have refused to accept justifications for the exclusion of osteopaths of the sort just suggested, thus evidencing an intermediate level of review. However, other courts have sustained discrimination against osteopaths with very little justification. *E.g.,* Stern v. Tarrant County Hosp. Dist., 778 F.2d 1052 (5th Cir.1985) (en banc)

(upholding exclusion of osteopaths, over a vigorous dissent that condemned "the bigotry of an allopathic-dominated state hospital district that refuses to be bothered by either the state law, the federal constitution or the facts").

Courts have also reached conflicting results over whether a hospital can discriminate on the basis of medical certification. In contrast with excluding incompetent doctors, a hospital may want to limit staff membership to only the most highly qualified doctors, as designated by "board certification," which some courts find permissible. The indirect effect of this requirement, however, may be to exclude certain schools of practitioners who do not meet the eligibility requirements, practitioners such as osteopaths. Therefore, other courts have struck down these policies as discriminatory.

The great bulk of case law addressing medical staff disputes concerns the required elements of *procedural* fairness for denying or revoking staff privileges, rather than the substance of those decisions. Much of this body of law has essentially the same content as general administrative law (even though private hospitals are not subject to the Administrative Procedure Act or the Constitution), so it will receive only cursory discussion here. This law is now largely determined by the federal statute discussed below which creates a qualified immunity for peer review participants if the process is conducted fairly. The statute and its regulations detail the procedural components necessary to qualify for this immunity, which essentially codifies

the existing state case law for common law fairness. Procedural steps are also usually specified to some extent in the medical staff bylaws.

Generally speaking, hospitals must provide physicians a formal evidentiary hearing before taking any final adverse action on either expulsion or admission. In the case of initial applicants, this hearing normally occurs as a review of the hospital board's initial decision to deny the application. In the case of revoked or suspended privileges, the hearing usually occurs once an initial decision is made to consider taking disciplinary action against the physician. Prior to the hearing, the hospital must give the doctor adequate notice of the charges and evidence against him. The doctor must be allowed to appear and present contrary evidence. Members of the hearing panel must not have had prior exposure to the case or otherwise have formed a biased or predetermined position on the merits. (The latter prohibition may be particularly troublesome in small rural hospitals, where it may be necessary to use a hearing panel composed of persons from outside the hospital.) In other respects, these hearings need not be conducted with the formality of a courtroom proceeding. For instance, most courts agree that lawyers can be excluded from the proceeding and that the hearing can be conducted without cross examination.

5. PHYSICIAN MEMBERSHIP IN MANAGED CARE NETWORKS

As noted above, these same issues exist with respect to physician membership in managed care networks such as HMOs, but the legal framework differs to some extent. At the outset, one might ask why insurers should have to engage in further credentialing,

considering that physicians are already screened by licensing authorities and several hospitals. As a practical matter, this debate is moot since insurers have decided to do this voluntarily, perhaps prompted by the threat of corporate liability, and credentialing is generally required by managed care licensing laws and accreditation standards. The issue then arises what legal protections physicians have when they are excluded.

The legal focus so far has been entirely on the removal of existing network members, which is known as "deselection," rather than on the denial of initial applicants. In part, this is because the available legal theories don't support initial applicants to the same degree as with hospitals, and in part because managed care networks are not viewed or established in the same open staff model as are community hospitals. But, this area of law is still in development.

First, consider whether the quasi-public facility characterization should apply to insurers. The case is much weaker than for hospitals, since HMOs are much less likely to be monopolistic, and there is a lesser degree of government support. Intuitively, the public does not regard them in the same light as a

community institution that should be open to every physician, nor do they usually hold themselves out in this manner. Therefore, most courts decline to extend the hospital precedents to HMOs. So far, only California differs. Potvin v. Metropolitan Life Ins. Co., 997 P.2d 1153 (Cal. 2000).

Other courts, however, have adopted a different legal theory for judicial scrutiny of HMO staffing decisions. They have applied the employment-at-will precedents (discussed above) to contract terminations that violate public policy. Where this cause of action

exists, how does the resulting level of scrutiny differ from under the quasi-public facility theory? First, note that the at-will contract theory applies only to existing network members, not to applicants. Second, note that the theory explicitly allows "no-cause" terminations. The burden is on the excluded physician to show there was a hidden bad cause that is contrary to public policy. Reasons that are merely arbitrary or poorly supported, which would violate quasi-public review, are permissible under this theory. What does in fact violate public policy? Retaliating against physicians who stand up for patients' rights, similar to the "whistleblower" cases discussed below, is a prime example. But, it would appear that economic grounds or personality disputes, even if manifestly unfair, are permissible because they don't contravene established public policy.

Finally, as for process rights, managed care network physicians, like hospital staff members, have a right to receive the process that is promised in

their contracts, but, as noted above, managed care contracts frequently contain no-cause termination provisions that allow immediate dismissal for no stated reason.

6. PEER REVIEW CONFIDENTIALITY AND IMMUNITY

Courts adjudicating medical staff disputes have been reluctant to interfere in the internal workings of a hospital or HMO, both because they feel unqualified to make judgments about medical competence and because they fear that such litigation will deter physicians from performing their vital peer review role effectively. Intensive peer review is considered essential to improving the quality of medical care, a matter of heightened concern as a result

of the medical malpractice "crisis." Therefore, a number of state and federal enactments clothe the medical peer review process with confidentiality or immunity.

a. State Peer Review Confidentiality Statutes

Some states have "shield laws" that provide a degree of legal immunity for hospital peer-review committee members and witnesses by protecting them, for instance, from defamation suits. A greater number of states have a peer review *confidentiality* statute that protects the proceedings of any hospital peer review committee from discovery, and a few decisions have found a common-law basis for such a privilege even in the absence of such statutes. This peer review privilege shields internal hospital review

records both in actions *by* the sanctioned doctor (for improper exclusion) and in actions *against* the doctor (for malpractice). However, these statutes prohibit disclosure only of information and reports that are generated by internal review committees; they do not entirely remove relevant information from the litigation process if the information is derived from independent, noncommittee sources.

b. Federal Peer Review Immunity

The federal Health Care Quality Improvement Act of 1986, 42 U.S.C. § 11101 *et seq.*, confers sweeping peer review immunity, but of a highly qualified sort. On first reading, this Act would seem to render irrelevant much of the discussion in this and the next chapter, for it states that "any person who participates with . . . [medical peer review] action shall not be liable in damages under any law of the United States or of any State" except for civil rights

laws. However, there are many important exclusions from and limitations to this immunity.

First, this federal peer review immunity applies only to decisions involving physicians; it does not shield decisions to exclude other medical professionals. Second, even for physicians, the law is concerned only with the exclusion of individual physicians; immunity does not attach to the establishment of general credentialing criteria that exclude groups of physicians. Third, even for individual physician exclusions, the Act covers only exclusions based on conduct that could adversely affect the health of a patient, omitting decisions

based on economics or on more general ethical concerns or the welfare of the institution.

Even for those cases that do meet the Act's definition of "professional review action," immunity attaches only if the exclusion decision "meets all the standards specified [in the Act]." These detailed procedural and substantive standards in essence track much of the law that the Act attempts to preempt. To qualify for immunity, a peer review action must be taken:

> (1) in the reasonable belief that the action was in the furtherance of quality health care, (2) after a reasonable effort to obtain the facts of the matter, (3) after adequate notice and hearing procedures are afforded to the physician involved or after such other procedures as are fair to the physician under the circumstances, and (4) in the reasonable belief that the action was warranted by the facts.

42 U.S.C. § 11112.

Because of these limitations, many of the medical staff cases mentioned in this chapter and in Chapter 4.B (antitrust boycott law) would fail to qualify for immunity under the Act. Even for those that would, the impact of the Act is less than one might first think. A physician challenging an exclusion will still allege that the hospital's or insurer's action was not in good faith, was not supported by substantial evidence, and did not comport with fair procedures, all of which allegations, if true, would disqualify the defendants from the immunity. Thus, the

substantive core of these disputes has simply been shifted to the outset of the case. However, this shift has the important procedural effect of lessening the defendants' burden of proof: "A professional review action shall be presumed to have met the [required] standards . . . unless the presumption is rebutted by a preponderance of the evidence." *Id.* The statute also contains a fee shifting provision that penalizes frivolous challenges to peer review decisions. Finally, it is important to note that the statute immunizes only damages actions, not actions for injunctive relief.

C. LABOR AND EMPLOYMENT LAW

Labor and employment law is an area of emerging significance to health care institutions, not only because these topics are important to all industries, but also because these topics contain dimensions that are unique to the health care setting. Unlike most industries where the interests at stake are simply the employee's right to a job and the employer's right to do business as it sees fit, here the courts must consider how the law should be shaped to accommodate the additional concerns of avoiding disruption in medical services. Moreover, precedents drawn from conventional employment

contexts are sometimes not easily transferrable to the independent character of licensed medical professionals.

178

1. LABOR LAW

a. Patient Care Concerns

The law of collective bargaining provides the clearest examples of the unique problems that medical employment raises. The National Labor Relations Act ("NLRA"), 29 U.S.C. § 151 et seq., establishes a comprehensive regulatory scheme, administered by the National Labor Relations Board ("NLRB"), to protect employees' rights (1) to form a union, (2) to require their employer to bargain with the union in good faith and honor collective bargaining agreements, and (3) to strike if they are dissatisfied with the terms or conditions of their employment. Prior to 1974, nonprofit hospitals were not subject to this Act, but in that year Congress added provisions directed specifically to the health care industry. In doing so, Congress expressed a concern that union activity not disrupt hospitals' important patient care mission. Specifically, both the Senate and the House expressed concern that too many unions within a single institution would pose an excessive risk of crippling, repetitive strikes: "Due consideration should be given by the Board to preventing proliferation of bargaining units in the health care industry." 1974 U.S.Code Cong. & Ad.News 3950.

Incorporating this "nonproliferation" mandate into labor law has proven troublesome. The controversy turns on how the NLRB should frame its test for determining the proper definition of a hospital bargaining unit. Normally, the NLRB makes unit

determinations according to a "community of interests" test, that is, whether there is such a close community of interests between, say, RNs and LPNs to include them both in the same union. However, in response to a series of reversals holding that this test did not give sufficient weight to the nonproliferation mandate, the NLRB, which is notorious for deciding all policy issues in adjudication, took the extraordinary step of initiating a rulemaking proceeding to define the proper approach to health care unit determinations. Its final regulations allow up to eight separate bargaining units: RNs, physicians, other professionals, technicians, clerical, maintenance, security guards, and others. 29 C.F.R. § 103.30, 54 Fed.Reg. 16336 (April 21, 1989). These rules generated intense opposition from health care management, which perceives that carving units into smaller divisions greatly increases the ease of unionizing and therefore is inconsistent with the nonproliferation policy.

b. Physician Unions

A question of labor law that has gained importance is whether the NLRA protects physician unions. In order to resist mounting institutional and economic pressures, some doctors have attempted to form unions, particularly in the context of managed care contracting. However, limitations in the NLRA's coverage place substantial obstacles in their path. First, only employees have the right to unionize. This limitation has been used to exclude medical residents and interns from the protection of the Act under the theory that they are serving primarily as students,

not employees. The limitation to employees would also seem to exclude independent physicians on a hospital medical staff from the NLRA's protections.

Second, even doctors who are employed may not be covered because the Act has long been interpreted to exclude "managerial employees" from its protections—those employees who assist management in determining and implementing policy. NLRB v. Bell Aerospace, 416 U.S. 267 (1974). This exclusion creates the prospect that employed physicians are never covered by the Act because health care institutions universally require heavy physician involvement on numerous committees that establish important medical policy. In a similar context, the Supreme Court determined (by a 5–4 vote) that the managerial employee exclusion applies to university professors because they participate on important committees and heavily influence faculty hiring decisions. NLRB v. Yeshiva University, 444 U.S. 672 (1980). However, the *Yeshiva* Court was wary of excluding all professional employees by this decision. Therefore, it cautioned that "employees whose decisionmaking is limited to the routine discharge of professional duties in projects to which they have been assigned cannot be excluded from coverage. . . . Only if an employee's activities fall outside the scope of the duties routinely performed by similarly situated professionals will he be found aligned with management." 444 U.S. at 690.

The NLRB has since wrestled with what this ruling might mean for the various administrative duties typically assigned to physicians within

medical institutions, reaching conflicting results in its two principal decisions. In Montefiore Hospital, 261 N.L.R.B. 569 (1982), the Board distinguished *Yeshiva* in the context of employed physicians at a large teaching hospital, allowing them to form a union (even though, by coincidence, this hospital provided the teaching faculty for Yeshiva University's medical school!). The Board reasoned that, due to the highly structured hierarchical nature of the hospital's departmental organization, staff committees merely provided recommendations to departmental chairs and had no ultimate policymaking authority. However, in FHP, Inc., 274 N.L.R.B. 1141 (1985), the Board followed *Yeshiva* and rejected a physician union at an HMO, giving an extensive rendition of the various committees on which physicians served. The Board observed that "many of the decisions made at the committee level . . . lie at the core of the [HMO's] operations" and that "recommendations are regularly if not always followed." But later the Board reversed course again and held that physicians at a different HMO are not managerial and so can unionize. Thomas-Davis Medical Centers, 324 NLRB No. 15 (July 24, 1997).

Similar confusion reigns with respect to nurse managers. In 1994, the Supreme Court ruled that nurses employed in a nursing home exercised sufficient de facto supervision over the work assignments of nurses' aides to be subject to the managerial exclusion. NLRB v. Health Care & Retirement Corp. of America, 511 U.S. 571 (1994). The NLRB, however, has resisted following this precedent, for instance by finding that hospital

"charge nurses" do not have managerial status. In NLRB v. Kentucky River Community Care, 532 U.S. 706 (2001), the Court reversed the NLRB again, holding that it read the Act too narrowly

in ruling that supervising nurses at a mental health facility are not managers. Yet, the NLRB once again resisted, ruling that only full-time charge nurses qualify as managers, and not those who have these responsibilities only part time. Oakwood Healthcare Inc., 348 NLRB No. 37 (2006).

It is difficult to reconcile these various decisions, other than to observe that the inquiry is a highly fact-sensitive one in which the particulars of proof and the quality of lawyering makes a critical difference. In reality, it is unlikely that there is a tremendous difference between the authority of physician committees and nurse supervisors in these various settings. This unstable law results from the difficulty of adapting conventional notions of hierarchical corporate decisionmaking to the collegial environment of medical professionals.

2. EMPLOYMENT LAW

a. Wrongful Discharge

In addition to the public regulatory law just surveyed, there have been important developments in private law doctrine that affect employment relationships in the health care sector. The most prominent development is the abrogation of the employment-at-will doctrine. Historically, employers have been free to fire workers with or without cause

if they were hired without a specific contractual term of service. Over the last two decades, courts have so frequently found exceptions to this common law rule that some lawyers view the exceptions as having swallowed the rule. This doctrinal development is discussed here because, for some unknown reason,

these employment-at-will disputes arise with surprising frequency in the health care industry.

There are three possible avenues for qualifying an at-will employer's freedom to fire. First, courts may find that the reason for firing an employee violates public policy, such as where an employer fires a "whistleblower" who reports safety violations to the proper authorities. See the discussion in section B.3 of managed care "deselection" suits. Second, employee manuals sometimes promise that procedural or substantive safeguards attach to the decision to terminate. If so, courts may find that these manual provisions have become implied-in-fact terms of the employment contract, under the reasoning that employees relied on these promised protections in deciding to stay on the job. Third, courts may rely on the covenant of good faith and fair dealing that the law implies into all contracts to hold that employers must give some good reason, and some fair process, when dismissing an employee, but this is only a minority position in the states.

Contrary to appearances, then, there has been no general abrogation of at-will employment in most jurisdictions; the exceptions to this rule remain just that. If employee manuals are carefully drafted, employers can avoid the contractual theory. As for

the public policy theory, not every questionable reason for discharge violates public policy. Also, courts are careful to distinguish discharges based on behavior motivated by private concerns versus those motivated by public concerns. One situation, however, where health care employees are protected in their exercise of personal conscience is with respect to abortion. State

and federal "conscience clause" statutes protect health care employees who refuse to perform abortions. *E.g.* 42 U.S.C. § 300a–7.

b. Covenants Not to Compete

A final area of importance in health care employment law arises from covenants not to compete in physician employment or managed care contracts. In order to protect a practice group's interest in its established patients, physicians joining a group are usually required to abide by restrictive covenants that prevent them from practicing in the same geographic area for a certain period of time once they leave the group. However, courts view covenants not to compete in this and other employment contexts with suspicion because these restrictions contravene public policies in favor of free trade and the right to work. Restrictive covenants are valid only if they are reasonable as to time, geographic scope, and the range of activities covered, and if they are not otherwise contrary to the public interest. *See, e.g.,* Karpinski v. Ingrasci, 320 N.Y.S.2d 1, 268 N.E.2d 751 (1971) (restriction of defendant from practice of dentistry is overbroad where his employer practiced only oral surgery).

Of particular relevance in the medical context is whether it contravenes the public interest to restrain the practice of a physician whose services are needed in the community. According to some courts, "The right of a person to choose the physician that he or she believes is best able to provide treatment is so fundamental that we can not allow it to be denied because of an employer's restrictive covenant." Murfreesboro Medical Clinic v. Udom, 166 S.W.3d 674 (Tenn. 2005). Another line of cases has

reasoned, though, that the public interest is equally well served even if the doctor is forced to treat patients in another part of the state.

CHAPTER 4
ANTITRUST LAW AND HEALTH CARE

A. INTRODUCTION

The federal antitrust laws have prompted one of the most burgeoning areas of health care litigation in modern times. Prior to 1980, antitrust challenges to the health care industry were almost unheard of. It was thought that the antitrust laws did not apply to the learned professions and that health care is an inherently local activity not subject to federal jurisdiction. The Supreme Court exploded both of these myths in the mid-1970s. In Goldfarb v. Virginia State Bar, 421 U.S. 773 (1975), a case concerning the legal profession, the Court held that "the nature of an occupation, standing alone, does not provide sanctuary from the Sherman Act." A year later, the Court reinstated an antitrust suit against a hospital, ruling that its operations have a substantial impact on interstate commerce by virtue of the purchase of equipment and supplies and the receipt of insurance reimbursement from out of state. Hospital Building Company v. Trustees of Rex Hospital, 425 U.S. 738 (1976).

The two-fisted blow of these decisions has left the entire health care sector reeling ever since. Exposure to antitrust scrutiny is a frightening prospect to doctors and hospitals because of the enormous potential liability. Damages awarded for lost business can be substantial and, under the antitrust

laws, these damages are automatically tripled. Antitrust defendants may also have to pay for the costs of the plaintiff's attorneys if they lose. And, while most doctors and hospitals are well insured for malpractice, their insurance often does not cover antitrust liability. Finally, criminal enforcement actions can be brought by the government, possibly resulting in imprisonment.

This chapter explores four branches of health care antitrust law; group boycott challenges to a hospital's or HMO's exclusion of physicians from the medical staff; the price-fixing ramifications of health insurance and provider networks; vertical restraints and monopolization charges against HMOs; and merger doctrine as applied to various health care ventures.

B. ANTITRUST BOYCOTT LAW

Increasingly, medical staff disputes boil over into antitrust litigation, usually under section one of the Sherman Act, which prohibits conspiracies in restraint of interstate trade. 15 U.S.C. § 1. Physicians who are either denied admittance to the medical staff or whose staff privileges are revoked regularly characterize their exclusions as "concerted refusals to deal," or, more pejoratively, as group boycotts, which the Supreme Court has held from time to time in other contexts are per se illegal. Although antitrust law as a body is complex, this particular theory of action is rather straightforward; it is anticompetitive to allow physicians to exclude

their competitors from hospital facilities that are essential for the practice of medicine.

1. THE CONSPIRACY REQUIREMENT

In order for such a charge to be entertained under Sherman Act section one, the plaintiff must establish that a medical staff exclusion was the product of concerted (joint) rather than unilateral (single actor) behavior. This might seem to be a simple matter given the multiplicity of actors that usually participate in hospital credentialing decisions, but antitrust law generally looks to economic *entities* rather than individual *persons*. For example, it clearly does not constitute price fixing for the managers of a manufacturing firm to meet together to decide how to price their product, even though many people are involved in the decision, because they are all acting in the economic interest of a single entity: their employer. Such behavior does not present any anticompetitive threat because it does not "suddenly bring together economic power that was previously pursuing divergent goals." Copperweld Corp. v. Independence Tube Corp., 467 U.S. 752, 769 (1984). Likewise, in the context of medical staff membership decisions, it would initially appear that no conspiracy exists because physicians who participate in hospital peer review do so under the auspices of the hospital corporate entity.

However, there is an exception to this "intra-enterprise conspiracy" rule that is particularly problematic in this context. The rule applies only if those who claim its protection are pursuing a single

economic interest: that of the hospital. Physicians, though, are not hospital employees; they have their own independent practices. Therefore, it is more than possible that a given physician might act out of the personal economic motive of not losing patients rather than out of concern for the quality of care at the hospital.

Most courts view the independent interest exception to the intra-enterprise conspiracy rule as a highly fact-sensitive judgment that inquires into the subjective motivations of the defendants. However, one early and influential decision, Weiss v. York Hospital, 745 F.2d 786 (3rd Cir.1984), seemed to hold that all physicians who participate in peer review are engaged in a conspiracy *as a matter of law*! *Weiss* concerned the exclusion of an osteopath from a hospital medical staff controlled by M.D.s, and the court held that "the York medical staff is a group of doctors, all of whom practice medicine in their individual capacities, and each of whom is an independent economic entity in competition with other doctors in the York medical community."

There are two obvious difficulties with this analysis. First, it isn't true that all doctors have a competitive interest at stake with all other doctors. If Dr. Weiss (the physician who applied for privileges at York Hospital) had been a cardiac surgeon who limited his practice to pediatric patients, he might have been in competition with very few doctors on the medical staff. Even as a general practitioner, he was not in competition with anesthesiologists, radiologists or pathologists. Second, even where

potential for an independent, anticompetitive stake exists, it would seem unfair to presume as a matter of law that no physician is capable of setting that interest aside and acting solely out of the hospital's concern for quality of care. Therefore, other decisions have allowed factual presentations on these issues, but that hardly constitutes a safe harbor. Accordingly, the safest alternative is to structure the peer review process so that only non-competing doctors are involved, even if this means delegating review

functions to doctors who live elsewhere. Medical staffs typically are unwilling to relinquish this control, however.

2. MEDICAL STAFF BOYCOTTS AS RESTRAINTS OF TRADE

a. Avoiding Per Se Illegality

A Sherman Act section one plaintiff, if she clears the conspiracy hurdle, must next establish that the challenged behavior constitutes a restraint of trade. The courts have made a broad division between two types of tests to judge the multitude of economic behavior that potentially constitutes a restraint of trade. If a given practice is so obviously anticompetitive that it can virtually never be justified—for example, price fixing among competitors—the courts apply the "per se illegal" label and find an antitrust violation as a matter of law. All other challenged business activity is judged under some version of the rule of reason, which seeks to balance the activity's anticompetitive effects against its procompetitive effects.

This reasonableness inquiry traditionally has required a lengthy trial with a host of economic and industry experts. Therefore, even this lesser standard of liability gives the physician-plaintiff strong leverage for bargaining since hospitals will find it difficult to prevail at a summary judgment stage under this fact-sensitive analysis. However, the federal courts have developed more lenient forms of the rule of reason that demand a greater showing of the plaintiff in order to survive the defendant's motion for summary judgment.

Group boycotts, otherwise known as "concerted refusals to deal," constitute one category of activity that has received the per se label.

In the paradigm case, a group of retailers refuses to deal with a wholesaler who is cooperating with another retailer outside of the group. But "group boycott" is simply a pejorative label that might attach to any concerted refusal to associate with someone else, and one can think of any number of such actions that are socially beneficial, such as where a professional society excludes a member for engaging in unethical behavior. Therefore, a great deal of controversy surrounds the general question in antitrust theory of which group boycotts are automatically per se illegal.

In the context of hospital peer review, Weiss v. York Hospital, 745 F.2d 786 (3rd Cir.1984), is again instructive. There, the court wrestled with fitting a medical staff exclusion decision into the classic boycott mold. A hospital might be viewed in the chain of production of medical services as equivalent to a wholesaler, and doctors might be viewed as retailers.

Therefore, to match the paradigm case, doctors would have to exclude a competitor by threatening to leave the hospital. This could be called a "secondary boycott" because it is directed at one party (the hospital) in order to reach a second (the competing physician). Doctors have little reason to engage in the classic paradigm of a secondary boycott, however, since they have the ability to engage directly in a *primary* boycott of an unwanted competitor. Existing medical staff members, because of their influence over credentialing decisions, can exclude a competitor simply by deciding themselves not to deal with a medical staff applicant rather than attempting to coerce the hospital not to deal with the competitor.

The *Weiss* court found this incongruity in paradigms much more troubling than it actually is. Primary boycotts are rare in antitrust precedents only because few industries are structured like the medical profession to allow existing competitors to expel or block the entry of potential competitors. The per se illegality of primary boycotts follows *a fortiori* from the precedents that address secondary boycotts.

Thus, there are strong precedential and theoretical grounds to be wary of medical staff exclusions. On the other hand, the role of physician peer review is vital to maintaining the safety of medical treatment. State medical licensure boards (like state bars) do little to ensure the continuing competence of practitioners, so the critical weeding-out task falls on hospitals, who must necessarily delegate the decision to those with knowledge and expertise—their member physicians.

Moreover, it is through the peer review process that hospitals compete among themselves on the basis of their quality of care by attempting to select the very best doctors available. To essentially outlaw medical peer review therefore would have disastrous consequences.

Medical peer review thus confronts antitrust analysis with a unique paradox; heightened anticompetitive risk coupled with heightened social need. As a result, this is not an area that lends itself well to the traditional dichotomy offered by the sharp doctrinal distinction between the rule of reason and per se illegality. What one would like to do is to take a quick look at each case to see if legitimate reasons have been advanced for adverse action

against a medical professional. If so, then the rule of reason would be appropriate. If not, then the per se rule should be applied.

This bifurcated analysis appears to reflect the dominant approach that the courts have in fact taken in medical staff cases. In *Weiss v. York Hospital*, *supra,* for instance, the jury found that Dr. Weiss had been excluded as a result of the medical staff's bias against osteopaths generally and the defendants did not attempt to advance a legitimate (*i.e.,* quality-of-care based) justification for their discriminatory action. Consequently, the court applied the per se label. However, the court stated that had the hospital excluded osteopaths "on the basis of their lack of professional competence or unprofessional conduct, . . . the rule of reason, rather than a per se rule, would be applicable."

195

This ruling appears to comport with the Supreme Court's subsequent decisions in Northwest Wholesale Stationers Inc. v. Pacific Stationery and Printing Co., 472 U.S. 284, 294 (1985) and in California Dental Ass'n v. FTC, 526 U.S. 756 (1999). In *Northwest Wholesale* the Court explained in another context that "the mere allegation of a concerted refusal to deal does not suffice because not all concerted refusals to deal are predominantly anticompetitive." Therefore, "a plaintiff seeking application of the *per se* rule must present a threshold case that the challenged activity falls into a category likely to have predominantly anticompetitive effects." The Court noted that, in cases where this showing has been made in the past, the challenged practices "generally were not justified by plausible arguments that they were intended to enhance overall efficiency and make markets more competitive." In *California Dental,* the Court ruled that a trade association ban on most forms of price and quality information in

advertising by dentists might be permissible, stating that it wanted to see a "less quick look" than the courts had given because each case requires "an enquiry meet for the case, looking to a restraint's circumstances, details, and logic."

In *California Dental*, the Court stressed that providers had made a serious argument that quality of care might suffer without the restraint in question. In contrast, the York Hospital medical staff's failure to articulate a quality-of-care justification for a medical staff membership policy is rare in antitrust litigation. Virtually every other case has been

decided under a rule of reason analysis because most hospitals readily establish at least the pretext of a quality justification. For example, in F.T.C. v. Indiana Federation of Dentists, 476 U.S. 447 (1986), the Court applied the rule of reason to a professional dentist society's refusal to submit dental X-rays to insurance companies, even though this is ordinarily a per se illegal group boycott, because the dentists justified this practice based on the quality-of-care rationale that X-rays alone are an insufficient basis on which to determine the medical necessity for dental services.

b. Avoiding Trial on the Merits

It is fairly easy for antitrust defenders to make the necessary allegations to avoid per se condemnation of medical staff decisions, but that is only the first line of defense in these cases. Defense attorneys are also eager to avoid trial on the merits, both because this is so time-consuming and expensive, and also because individual physicians denied the chance to practice their profession are sympathetic figures in jury trials. Courts too are reluctant to let these cases go to trial out of the judicial instinct that they are

nothing more than individual personnel disputes, which should not be turned into federal antitrust cases.

This instinct has expressed itself in a number of different areas of doctrine that allow courts to render summary judgment for the defense, despite plausible assertions of anticompetitive conduct by excluded physicians. Most of these rulings rely on the notion

that antitrust law protects competition rather than competitors. This slogan reflects the idea that, even though the economic harm to an individual physician may be great, competition itself is not harmed as long as there are plenty of other physicians around to pick up the slack. Thus, courts are inclined to dismiss these cases outright if the remaining physicians do not exercise significant market power. Even when courts allow these cases to go to trial, they frequently exert a very lenient level of scrutiny, one that asks only whether the hospital had a rational basis and some evidentiary support for the exclusion, and not whether, on the merits, the court thinks the charges against the physician are true. As advocated by Professor Clark Havighurst, a leading commentator on health care antitrust:

> To ensure that hospitals have reasonable freedom of action, . . . summary judgment or a directed verdict would be appropriate if documentary evidence and affidavits showed that the hospital's action reflected its [own] corporate concerns. . . .
>
> Under this test, a court would not concern itself . . . with whether the ostensible motives for the actions taken were the real motives or whether the adverse effect of the hospital action on competition among practitioners was outweighed by its actual contribution to fulfilling the hospital's objectives.

Clark Havighurst, *Doctors and Hospitals; An Antitrust Perspective on Traditional Relationships*, 1984 Duke L.J. 1071, 1133–34, 1157.

Courts are most likely to follow this course when adverse action is taken against a single physician based on his individual competence, rather than in a case that excludes an entire class of practitioners or that is clearly motivated by economic rivalry. The same attitude is captured in the qualified immunity statute discussed in section 4.C.6.b, which applies only to individual physician exclusions based on quality, and not to group-based exclusions, exclusions of non-physicians, or those based on economic grounds.

3. EXCLUSIONS FROM MANAGED CARE NETWORKS AND ILLEGAL TIE-INS

This same body of law potentially applies as well to the exclusion of physicians from managed care networks. One of the leading cases serves well to illustrate how the principles above apply in this slightly different context. In Hassan v. Independent Practice Associates, 698 F.Supp. 679 (E.D. Mich. 1988), an HMO removed from its network two allergists whom it thought were ordering too many tests. First, the court found there was a potential conspiracy since primary care physicians were involved in the decision and they compete for some of the same business. Next, however, the court declined to apply the per se rule of illegality. Even though no quality of care issues were raised, the economic issues advanced by the HMO appeared legitimate. When it came to examining the merits, however, the court ruled on summary judgment in favor of the HMO. It observed that the HMO had little market power since the HMO controlled only 20 percent of

199

the insurance market, and the allergists were free to contract with other plans. Also, the other physicians had nonexclusive contracts with the HMO which made it easier for new HMOs to enter the market. Finally, the HMO subsequently admitted two other allergists to the network, which contradicted the claim that primary care physicians were trying to take over the allergy business.

One way plaintiffs might respond to these arguments is that an insurer's market power is potentially much stronger if viewed more narrowly, by looking at the set of patients who are current subscribers. For them, a physician's exclusion from the network has dramatic effect because, as long as they are members, they cannot shop for care from excluded physicians. For this set of patients, the HMO has 100 percent market share, and options within the network are limited. This analysis resonates with Eastman Kodak v. Image Technical Services, 504 U.S. 451 (1992), in which the Court held that a separate submarket could exist for the parts required to repair a Kodak brand photocopy machine if Kodak parts are not interchangeable with parts from other brands, since once someone purchases the expensive machine, they are locked in to buying only Kodak parts. This is similar to another theory of antitrust injury known as an illegal tie-in, which has also been declared to be per se illegal. A tie-in exists when a firm with market power over one product insists that purchasers also pay for another product they don't necessarily want.

200

Whether this line of analysis is convincing is subject to debate. HMOs do have a lock-in, but not for life, only for potentially a year, and perhaps much shorter. Moreover, the lock-in is often not

absolute, since many plans now allow patients to go outside the network by paying a somewhat higher copay or deductible. As for the tie-in characterization, the Court surprisingly accepted this characterization in a case in which a hospital required surgery patients to use only the anesthesiologists it had under exclusive contract. However, the Court found no violation since the particular hospital lacked the requisite market power to coerce the patient's choice of an anesthesiologist. Jefferson Parish Hospital District v. Hyde, 466 U.S. 2 (1984).

It appears, then, that insurers have fairly free reign to pick and drop their network providers as they see fit. Whether courts would be as lenient with hospitals for their economic-based exclusions is another matter. One is used to condemning hospital exclusions if they are based on the economic motives of member *physicians,* but in today's market environment, hospitals have their *own* economic interests to protect, and these, in contrast with member physicians' economic interests, are a legitimate basis for an institutional decision. After all, hospitals' quality interest is, at bottom, economic in nature, at least in the antitrust mindset, since maintaining quality enhances one's reputation in the market and reduces malpractice liability.

4. THE PATIENT CARE DEFENSE

Some medical staff exclusion cases do receive close antitrust scrutiny because, even though the stated reasons for exclusion might be legitimate enough to avoid per se condemnation, since there is enough potential for anticompetitive harm to warrant fact-based rule of reason balancing. For hospitals and professional associations, the critical issue then becomes precisely how quality

of care concerns are incorporated into what normally is a purely economic inquiry into the procompetitive versus anticompetitive effects of a refusal to deal.

For a time, it was thought that all of the learned professions might impliedly be exempt from the antitrust laws because of the influence of professional ethics and public service norms on their behavior. But in Goldfarb v. Virginia State Bar, 421 U.S. 773 (1975), a case concerning lawyers' fee schedules, the Supreme Court rejected any such sweeping exemption. However, the Court left behind some pregnant dictum whose meaning still has not been fully articulated; "the public service aspect, and other features of the professions, may require that a particular practice, which could properly be viewed as a violation of the Sherman Act in another context, be treated differently." 421 U.S. at 778 n. 17.

One effect of this dictum appears to be that conduct otherwise falling within the per se prohibition will be judged under the rule of reason if the conduct is "premised on public service or ethical norms." Arizona v. Maricopa County Medical Society, 457 U.S. 332, 348–49 (1982). But this merely restates the

analysis above. It does not address the additional question of, assuming that the rule of reason applies, exactly how the rule takes account of these norms. At first, it was thought that the effect of the *Goldfarb* dictum would be to apply a sort of "relaxed" rule of reason analysis to the professions, one that gives heavy deference to judgments concerning the demands of public safety. However, the Court squarely rejected any such special approach for safety concerns in National Society of Professional Engineers v. United States, 435 U.S. 679 (1978). There, the Court held that the rule of

reason "does not open the field of antitrust inquiry to any argument in favor of a challenged restraint that may fall within the realm of reason. . . . [T]he inquiry is confined to a consideration of impact on competitive conditions." But later in California Dental Ass'n v. FTC, 526 U.S. 756 (1999), the Court reopened door, at least for medical professionals. There, the Court called for a level of scrutiny somewhere between per se and rule of reasons, to examine the claim that restraints on price and quality advertising by dentists are justified by "significant challenges to informed decisionmaking" about "the quality of professional services," which "tends to resist either calibration or monitoring by individual patients or clients."

The cloudy law on this question might appear to place hospitals and doctors in an intolerable dilemma: tort law forces them to review the competence of medical staff members, but antitrust law's opaqueness regarding non-economic justifications for physician exclusions appears to

make quality-of-care defenses legally risky. This appearance is misleading, however. It is patently wrong to conceive of economic motives and quality of care as mutually opposed considerations in hospital staffing decisions. Indeed, from the hospital's perspective, the two are *co*extensive because *quality is the key competitive variable in the hospital industry.* Indeed, for decades quality competition was virtually the sole force that drove the entire health care sector. Therefore, quality of care concerns factor directly into the procompetitive side of the rule of reason balance.

A good illustration of how this outline analysis might apply in an actual case comes from Wilk v. American Medical Association,

671 F.Supp. 1465 (N.D.Ill.1987), which addresses the AMA's long-standing opposition to the chiropractic profession. Although this case does not concern a medical staff exclusion, it reflects another body of caselaw that addresses rules set by professional societies. The district court found that the AMA's refusal to allow physicians to associate with chiropractors constitutes an unreasonable restraint of trade. In doing so, however, the court, following an earlier 7th Circuit opinion in the same case, allowed the AMA to assert an affirmative defense based on its concerns over the unscientific nature of chiropractic. This defense required the AMA to show that its concern was objectively reasonable and that it could not have been satisfied in less restrictive means than a total boycott. The court ruled that the AMA satisfied the first element but not the second.

Although *Wilk* reached the correct outcome, its reasoning has been criticized for misconstruing how quality of care concerns should be considered in more typical cases. Consider, for instance, a hospital that decides to exclude chiropractors. It seems excessive to place the burden of proof on the hospital to prove they are dangerous to patients. On the other hand, in the *Wilk* case, it appears inconsistent with Supreme Court precedent to potentially allow the AMA off the hook for its good motives if in fact the blatant boycott was anticompetitive. How does one resolve these conflicting impressions? The key is to distinguish between an ordinary defense and an affirmative defense. Where quality is legitimately a competitive factor for a single entity, as it is for a hospital but not for the AMA (which is a walking conspiracy), then quality of care should be considered as a direct or ordinary defense. An ordinary defense leaves the burden of proof on the plaintiff and

requires only a showing that the decision improved quality to some degree, not that it was necessary to avoid incompetence in any absolute sense. In this situation, an affirmative defense is unnecessary, and its proof standards are too demanding. However, where quality is not a legitimate competitive factor, then there is no justification for constructing a special defense that forgives antitrust violators because of their good intentions.

5. THE INTERSTATE COMMERCE AND STATE ACTION DEFENSES

Two other defenses deserve brief mention. For a time some courts rejected federal antitrust

challenges to medical disputes because they viewed medical care as an inherently local matter that does not constitute interstate commerce. However, the Supreme Court has clarified that there is sufficient impact on interstate commerce to invoke federal jurisdiction because medical care uses supplies from out of state and is often paid for by out-of-state insurers. Summit Health v. Pinhas, 500 U.S. 322 (1991).

A few other courts also rejected these challenges under a more obscure defense known as "state action immunity," which reasons that Congress never intended to apply antitrust law to state government. For instance, it would appear spurious to charge a state Board of Medical Examiners with an antitrust violation for revoking a physician's license based on incompetence. The state action defense potentially exists with respect to physician discipline even at *private* hospitals if state law mandates and "actively supervises" the private peer review process. The Court rejected this contention, however, in Patrick v. Burget, 486 U.S. 94

(1988), holding that private hospital credentialing decisions in Oregon are not sufficiently supervised by the state to bring them within the protection of this defense. The court observed that Oregon, unlike other states, had not yet recognized a common law right of judicial review of credentialing decisions.

Although *Patrick* was initially read as closing the door to state action immunity, it left open the small crack of a possibility that immunity might exist in another state whose common law differs. However, it

is unlikely that any common law theory of judicial review would meet the Court's objection that "the review [is] of a very limited nature . . . [in which the courts do no] more than make sure that some sort of reasonable procedure was afforded and that there was evidence from which it could be found that plaintiff's conduct posed a threat to patient care." On the other hand, a state might, if it chose to, impose this type of supervision through regulatory rather than judicial channels.

A number of states have done so in the form of "certificate of public advantage" (COPA) laws, which set up an explicit system for giving state regulatory blessing to cooperative agreements that are viewed as beneficial to the community but which might otherwise violate the antitrust laws. This state review process is designed to meet the active supervision requirement. These statutes do not apply to peer review decisions. Instead, they are directed to joint ventures and mergers, like those discussed below, in which hospitals and other providers seek to cooperate in order to achieve efficiencies and avoid costly duplication of facilities. For a time, it was thought such arrangements clearly met the *Patrick*. test. Some

doubt was created, however, by North Carolina State Board of Dental Examiners v. Federal Trade Commission, 574 U.S. 494 (2015), which held that even an official licensing authority is not immune from a policy that inhibited competition over teeth whitening, absent a showing that the state legislature actively supervised the Board or endorsed its teeth whitening policy. That decision appears to expose licensing boards to potential antitrust

scrutiny for a wide range of policies that might lie outside their core explicit missions.

C. PRICE FIXING AND VERTICAL RESTRAINTS INVOLVING INSURERS

1. PRICE FIXING IN PROVIDER NETWORKS

The primary example of per se illegality is horizontal price fixing, defined as any agreement among horizontal competitors (*i.e.,* firms at the same level within the marketplace) concerning the price or quality of their products, usually, an agreement not to undercut each others' price. The per se illegality of price fixing presents a danger in the establishment of various innovative forms of health insurance. The clearest example comes from preferred provider organizations (PPOs), but this analysis applies to other types of provider networks that enter into managed care arrangements. PPOs are groups of health care providers that agree to sell their services at a discount in exchange for receiving a large supply of business, say, all the employees covered by a large group insurance policy. PPOs are sometimes initiated by insurance companies or employers, who contact health care providers individually and negotiate discounts one by one. Such "consumer-based" PPOs raise

no price fixing concern because the consumer group establishes a separate bilateral contract with each provider. More typically, though, PPOs are initiated by doctors or hospitals who approach consumer groups and offer their discounted terms in a united fashion. (Perhaps this is more common because of the lower transaction costs

involved in this form of contract negotiation.) Provider-based PPOs must be organized as *groups* of doctors because no one physician could provide the bulk services required by a large insurance group. However, such "provider-based PPOs," and other kinds of provider networks, raise the distinct aura of price fixing since all of the doctors or hospitals first agree among themselves what price to charge or what discount to give to employers or insurers.

The price-fixing potential of provider-based PPOs was confirmed by the Supreme Court in Arizona v. Maricopa County Medical Society, 457 U.S. 332 (1982). Although *Maricopa County* did not address a PPO in name, the health care delivery plan at issue there (called a "foundation for medical care") was structurally identical to a PPO. The medical society in Maricopa County (Phoenix) sponsored a plan whereby all participating physicians agreed to abide by a maximum fee schedule if insurers agreed to pay 100% of those fees. Almost three quarters of the physicians in the county were members. The Court ruled that this arrangement constituted per se illegal price fixing. The Court was not concerned that the agreement set a cap on fees rather than a floor, since either could equally well serve as a benchmark for uniform pricing behavior, and, in any event, a cap on fees reduces *quality* competition if not *price* competition.

Maricopa County was decided by a 4–3 plurality vote as the result of two Justices abstaining. This fact, coupled with the subsequent change in Court members, raises the possibility that a majority of the

Court would reach a different result if confronted with this issue again, or would limit this holding strictly to its facts. The frailty of the *Maricopa County* holding is highlighted by a set of antitrust enforcement guidelines issued jointly by the Department of Justice Antitrust Division and the FTC. These guidelines do not directly alter the law, but they do state when these agencies will exercise their enforcement discretion to challenge provider-based PPOs and other contracting networks, such as accountable care organizations (ACOs). The DOJ/FTC guidelines in essence judge PPOs and ACOs under a rule of reason rather than a per se analysis. They express concern only if the arrangement involves a large percentage of the area's providers or if there is an overt indication of anticompetitive intent. The enforcement guidelines also establish several safe harbors that protect provider networks from any significant agency scrutiny. Therefore, despite the fact that *Maricopa County* still expresses established law, it is possible for thoughtful health care lawyers to devise several ways to sidestep the four corners of the *Maricopa County* holding by fine tuning the structure of provider-based PPOs to avoid application of the ominous price fixing characterization.

The cleanest way to avoid a horizontal price-fixing charge is to realign the agreement from a horizontal to a vertical dimension. This most clearly happens if the purchaser (insurer or employer) determines the price it is willing to pay and lets each doctor decide

unilaterally whether or not to join. Although this appears to produce exactly the same result—a group

of doctors agreeing on the same payment terms with one or more purchasers—the absence of a horizontal agreement among physicians removes this from per se price fixing.

The trouble is that physicians often want to take the initiative in forming these network arrangements, for both strategic and efficiency reasons. To allow this to continue, crafty health care lawyers have devised a way to dress a horizontal arrangement in vertical clothing, known as the "messenger model." In this arrangement, physicians agree to use a common negotiating agent to establish bilateral price agreements with purchasers. In the cleanest version, the agent conveys to each physician the insurer's price, and each physician decides whether to opt in. The DOJ/FTC guidelines approve this opt in version of the messenger model. Much more suspect, however, is a "black box" or "opt out" version in which physicians agree in advance to be bound by whatever payment terms the agent is able to negotiate, unless they opt out. Although physicians do not strictly speaking agree in advance to fixed-price terms, their pricing is sufficiently coordinated that the effect is virtually the same, so the enforcement guidelines prohibit this approach. Other inventive messenger approaches that fall between these two versions are also possible but have not been clearly ruled on.

A second technique for avoiding the *Maricopa County* holding is for the participating physicians, rather than merely establishing a loose contractual relationship, to integrate into a single economic

entity by forming, say, a large partnership in which they pool their assets and share business risks. Here again, inventive lawyering has found several ways to accomplish the substance or appearance of business integration without having to resort to the full scale version. The economic substance of integration can be achieved by exposing physicians to shared financial risk. The *Maricopa County* decision itself says this occurs if physicians accept capitation payment. The DOJ/FTC guidelines also mention fee withholds and other risk-based payment methods. The unresolved issues here are how extensive the financial risk must be, and whether the risk-based payments must cover all physician services or only some (primary care capitation versus global capitation). Note also that encouraging physicians to accept capitation payment to avoid antitrust risk conflicts with other legal signals discussed elsewhere, which discourage risk-based payment either by prohibiting or limiting financial conflicts of interest or by treating risk-bearing providers as if they were insurance companies.

A third possible way to avoid the *Maricopa County* holding is to create some new product. This avenue is suggested by the controversial and difficult case of Broadcast Music, Inc. v. Columbia Broadcasting System, Inc., 441 U.S. 1 (1979), where the Court held that the bulk licensing of music compositions does not constitute per se price fixing even though competing composers necessarily agree among themselves how much to charge for their songs. The Court reasoned in *BMI*, that this arrangement in effect creates a new product: a "blanket license" for

all BMI compositions, and price-setting is merely ancillary to this procompetitive device. Likewise, it might be contended that the bulk sale of medical services through a physician network creates a new product, particularly since networks usually offer a package of related services such as claims processing and utilization review. However, in *Maricopa County,* the Court distinguished *BMI* by holding in essence that these additional services amount to nothing more than window dressing and that PPOs do not offer a new and distinct health care product. The Court suggested that an HMO would meet this new product test, but this is simply another way of restating the financial integration defense just summarized.

The DOJ/FTC guidelines take a more lenient view, however, by holding out the possibility of avoiding per se condemnation by showing *clinical* integration. It appears that clinical integration includes looser arrangements than an actual group of physicians who practice in the same clinical setting. Instead, the guidelines suggest, and the FTC has upheld, the very same elements of peer review and common clinical protocols that were found insufficient in *Maricopa County*. This leaves considerable room for imaginative lawyers and managers to devise arrangements that will meet the requirements for a *BMI*-type defense.

A word of caution: most of what has been said in this section applies only to *physician*, not hospital, networks. Hospital networks raise much more serious antitrust concerns since there are far fewer

hospitals than doctors in any given market and so the dangers of collaboration are much greater.

2. MONOPOLIZATION BY INSURERS

More conventional insurance practices have also been subject to antitrust attack. Most cases concern insurers either limiting the amount that physicians can bill or excluding certain categories of health care providers. In one leading case, Blue Shield was sued unsuccessfully for imposing a ban on "balance billing," that is, prohibiting participating physicians from charging patients any more than the contractual amount paid by the insurance policy. Kartell v. Blue Shield of Massachusetts, 749 F.2d 922 (1st Cir.1984). In a second leading case, Blue Cross was sued successfully for limiting coverage of mental health services from clinical psychologists that were fully covered when rendered by psychiatrists. Virginia Academy of Clinical Psychologists v. Blue Shield, 624 F.2d 476 (4th Cir.1980). How can these two cases be reconciled?

Restrictive reimbursement policies are clearly valid under Sherman Act section one if they are imposed unilaterally by a single insurance company. This is a vertical, not a horizontal restraint, in which a purchaser simply declares how much it is willing to pay. Even monopolistic purchasers may do this. However, the historical fact that Blue Cross and Blue Shield are creatures of the hospital industry and the medical profession (respectively) suggests that it is sometimes possible to establish that insurance restrictions are the result of collaboration among

providers. For example, in Reazin v. Blue Cross & Blue Shield of Kansas, 899 F.2d 951 (10th Cir. 1990), the court sustained a $7.8 million jury verdict against Blue Cross for conspiring with one local hospital to exclude another from a PPO arrangement. In *Virginia Academy supra,* the court held that the mere fact that physicians controlled the board of Blue Shield was sufficient to

bring that company's unilateral actions under the purview of Sherman Act section one. However, most Blues plans have reconstituted their boards so that they are not dominated by physicians or hospitals. Consequently, it is now more difficult to show the existence of a horizontal conspiracy behind the actions of a single insurer and so more recent decisions have refused to entertain a section one theory.

Nevertheless, insurance restrictions may be challenged under Sherman Act section two, which prohibits unilateral monopolization or attempts to monopolize. The elements of these theories are not easy to meet, however. Monopolization requires a showing of a dominant market share and the use of illicit business practices. Attempted monopolization requires a showing of specific intent to monopolize and a likelihood of success. Most section two attacks have failed to meet these requirements. *See, e.g., Kartell supra* (no section two violation in limiting amount doctors can charge, despite 74% market share); Ball Memorial Hosp. v. Mutual Hosp. Ins. Co., 784 F.2d 1325 (7th Cir.1986) (no monopoly despite 50–80% market share).

Particular insurer practices have come under scrutiny, however. One is requiring that providers contract with a dominant insurer on an exclusive basis. This makes it very difficult for new insurers to enter the market, since existing providers have to give up their existing business to sign on with the new insurer. Therefore, the enforcement agencies and several court opinions look with disfavor on these exclusive provider contracts. One might respond that mutually exclusive competing provider networks are a superior market configuration than nonexclusive overlapping provider networks because this makes each insurers' product more distinct

and it allows insurers to impose more market discipline on providers by forcing them to engage in competitive bidding for network slots. These competitive benefits have to be weighed against the competitive harms of exclusive contracts, however, and most analysts believe that such contracts are dangerous in highly concentrated insurance or provider markets.

A related insurer technique is known as a "most favored nation" provision. In this technique, an existing dominant insurer (say, Blue Cross) tells its providers that they must give it the same payment discounts they agree to give any other insurer. This makes it very difficult for new insurers to enter the market by undercutting the dominant insurers' reimbursement rates, since any provider who agrees to the discounted rate for a small portion of their patients must then give the same discount for all their patients. Nevertheless, courts so far have upheld most favored nation provisions, reasoning that they are simply a way of requiring providers to

make their best price available to everyone. In the words of one court, "this, it would seem, is what competition should be all about." Ocean State Physicians Health Plan v. Blue Cross & Blue Shield of Rhode Island, 883 F.2d 1101 (1st Cir. 1989). Others, including the enforcement agencies, view this thinking as naive, especially in markets where an insurer has a dominant share, since this is simply another way to make market entry or expansion more difficult for potential competitors, similar to the use of exclusive contracts.

3. McCARRAN-FERGUSON EXEMPTION

Before leaving the arena of insurance and antitrust, it is necessary to observe briefly that insurance companies have a potential defense under the McCarran-Ferguson Act, 15 U.S.C. § 1011, which exempts the business of insurance from antitrust and other federal laws. Although this exemption appears sweeping, the Supreme Court has construed it very narrowly to apply only to those specific activities that are at the very core of insurance's risk-spreading function and that implicate the direct relationship between insurers and subscribers. *See* Chapter 5.C. Thus, the exemption does not apply to insurers' contracts or dealings with providers, since this is ancillary to or behind the scenes of the main insurance contract. Group Life & Health Insurance Co. v. Royal Drug Co., 440 U.S. 205 (1979).

D. MERGER LAW

Section seven of the Clayton Act prohibits mergers that "substantially lessen competition or tend to create a monopoly." 15 U.S.C. § 18. The principal concern of this prohibition is to keep markets from becoming so concentrated in a few large firms that it becomes easy for the major firms to collude. The legality of mergers is usually assessed under certain numerical tests that focus on the size and the number of firms and their market shares before and after a merger. Market share definition is also critical to other antitrust theories, so much of this discussion is relevant throughout this chapter.

One early hospital merger case is Hospital Corporation of America (HCA) v. Federal Trade Commission (FTC), 807 F.2d 1381 (7th Cir.1986), where Judge Posner wrote the opinion sustaining an FTC decision that prohibited the country's largest

hospital chain from acquiring several hospitals in the Chattanooga market. Prior to the acquisition, HCA owned only one hospital in the market. Afterwards, it owned or managed 5 of the 11 area hospitals. This acquisition resulted in four firms controlling 91 percent of the market. The court reasoned that this degree of market concentration is dangerous given the history of cooperation between competing hospitals in Chattanooga and the barrier that CON regulation creates to new hospitals entering the market.

Hospitals are not the only object of merger analysis in health care. Physician groups and managed care networks are also subject to merger scrutiny. The

DOJ/FTC guidelines set safe harbor protections that allow only 20–30 percent of the physicians in any given specialty to form a network without requiring agency approval. (The higher threshold applies only if physicians contract on a nonexclusive basis.) Similarly, a hospital's joint venture with its medical staff, for instance to operate an ambulatory surgery clinic, might be the subject of a merger challenge.

Because market share statistics control merger analysis, merger cases are often won or lost based on how the relevant market is defined. Ordinarily, the larger the market, the less significant will be the challenged firm's acquisition. There are two dimensions to the determination of market size; the geographic market and the product market. In the *HCA* case, for instance, the FTC used as the geographic market an area comprising the counties that are commonly recognized as encompassing the Chattanooga metropolitan area; the FTC rejected the contention that outlying rural counties should be included even though some patients from

those counties may occasionally travel into Chattanooga for highly specialized care. Other courts, however, have been convinced to accept geographic markets encompassing a dozen or so counties surrounding a metropolitan area, reasoning that, if prices increase too much, people will be willing to travel to receive care. One court observed that, even if this was not the case earlier, it is now very much the reality for HMO patients, since HMOs regularly force people to bypass nearby hospitals in order to use ones that the HMO contracts with. FTC v. Freeman Hospital, 69 F.3d 260 (8th Cir. 1995).

219

The other dimension of market definition is to define the relevant product. In *HCA,* the FTC looked solely at the operations of general hospitals. Thus, it excluded specialty hospitals that render only psychiatric care even though some general hospitals also treat psychiatric patients, but it included *all* the operations of general hospitals, both their inpatient business and their outpatient business, despite the fact that many nonhospital institutions also render similar outpatient care. These findings are recited not as stating binding precedent—for such determinations are highly fact-specific and thus are subject to proof in each case—but instead are given to illustrate the complexity of analysis that is required to determine whether a proposed merger creates an antitrust risk.

Consider also what the proper product market definition is in cases involving HMOs. In Blue Cross & Blue Shield United of Wisconsin v. Marshfield Clinic, 65 F.3d 1406 (7th Cir. 1995), another opinion by Judge Posner overturned a plaintiff's jury verdict based on a ruling that HMOs are not a separate market since they compete with indemnity insurers. This suggests markets for insurance whose scope is statewide or perhaps even national,

since the leading indemnity insurers compete nationally. The *Marshfield* case also discusses what the proper product scope is for physician markets. The court rejected the plaintiffs' allegation that physician services should be broken into markets for each specific service by DRG category depending on how many physicians are currently performing each procedure. The court observed that most medical

services can be performed by physicians in several different specialties so that if a given procedure were being monopolized there are always plenty of other physicians ready to compete. Therefore, Judge Posner appeared to use a market definition encompassing all physicians. The DOJ/FTC guidelines, in contrast, require that physician services be evaluated by specialty. They do not define, however, what constitutes a specialty. Consider, for instance, whether pediatric cardiac surgery constitutes a unique market or instead falls within the broader categories of surgery or cardiac surgery.

Although market share statistics dominate merger analysis, they are not the sole determinants of legality. Even mergers that monopolize a market may potentially pass muster: (1) if the acquired firm would have failed but for the merger; (2) if the market is too small to support more than one or a few hospitals (which is called a "natural monopoly"); or (3) if there are other factors that indicate the merger is beneficial to the public. For instance, the DOJ/FTC guidelines create a safe harbor for mergers involving hospitals with fewer than 100 beds and 40 patients a day, even if the merger results in only one hospital remaining in the market. Several courts, however, have allowed hospital mergers to go forward in highly concentrated markets because of the offsetting

concentration of buying power in insurers. In FTC v. Butterworth Health Corp., 946 F.Supp. 1285 (W.D. Mich. 1996), *aff'd*, 121 F.3d 708 (6th Cir. 1997), the court reversed the FTC and permitted the merger of the two dominant hospitals in Grand Rapids,

Michigan, observing that financially sound hospitals are necessary to counteract the growing threat of HMOs and "to continue the quest for establishment of world-class health facilities in West Michigan."

Initially, most health care merger law focused on hospitals rather than physicians. More recently, however, consolidation in the physician market is receiving increased antitrust attention. This consolidation is occurring in two significant ways beyond merely physicians expanding the size of their practice groups. First, hospitals increasingly are purchasing physician practices, to form larger, vertically-integrated delivery systems. Second, even when physicians remain independent, they are affiliating with each others and with hospitals more actively in order to form, via contract, "accountable care organizations." So far, there has been only limited litigation over the principles that govern these various physician affiliations, but to date those decisions have analyzed them under similar principles that apply to hospitals. See, e.g., St. Alphonsus Med. Ctr. v. St. Luke's Health Sys., 778 F.3d 775 (9th Cir. 2015).

A final area of controversy is whether the non-profit status of most hospitals alters the nature of merger (or other antitrust) analysis. There are two distinct issues. One is whether, as a technical jurisdictional matter, the Clayton Act applies at all to nonprofit institutions. This question has been resolved in the

affirmative. The second question is whether nevertheless there is less concern about mergers among nonprofits, following the logic that

nonprofits are less interested in profiteering by taking advantage of market power. Some cynics, such as Judge Posner in the *HCA* case and his other opinions, think nonprofits have essentially the same motive as for-profits to make as much money as possible. The only difference is what nonprofits can do with the money they make. Others, however, like *Butterworth supra*, believe that nonprofits are more likely to act cooperatively in a fashion that benefits the community at large by reducing excess capacity and improving coordination of services, following the health planning tradition reflected in the CON laws. This debate is informed to some extent by conflicting empirical findings, some of which show that mergers among nonprofit hospitals, even in highly concentrated markets, lead to reduced costs and prices, but others showing the contrary.

Each of these contested issues reveals that courts are taking an active role in setting antitrust policy for the health care industry. By doing so, courts for a time were much less deferential to the FTC and DOJ than they typically are in antitrust matters. During the 1990s and 2000s, there were repeated expressions of judicial annoyance at treating health care just like any other industry, and insistence that the enforcement agencies recognize the unique attributes of health insurance and medical care rather than rely on general economic theories. More recently, however, courts have started to recognize the competitive harms that may be resulting from consolidation in the health care industry.

CHAPTER 5

COMPLEX TRANSACTIONS AND ORGANIZATIONAL FORMS

As a consequence of increasing regulation and rapidly changing economic and legal forces, hospitals, doctors, and insurance companies have formed all manner of new organizational approaches to health care delivery and insurance, resulting in an explosion of acronyms. In the 1970s, we saw the birth of Health Maintenance Organizations (HMOs), a form of prepaid group medical practice explored in Chapter 1.A.2.f. This was followed in the 1980s by the creation of Preferred Provider Organizations (PPOs), a form of bulk purchase of medical services explored in Chapter 4.C.1. These are only the principal examples of integrated delivery systems ("IDSs"). Physicians and hospitals have formed joint venture organizations (sometimes called PHOs for physician-hospital organizations) for various purposes, and are now exploring new arrangements known as accountable care organizations (ACOs). Doctors have organized into large multi-specialty group practices known as independent practice associations (IPAs). And insurers of all types are using innovative methods (some of which are thankfully still unabbreviated) for controlling medical expenditures.

The result of this cauldron of activity has been to make the health care lawyer's work much more demanding—and in demand. Each of these innovative arrangements must be evaluated under a broad spectrum of private and public law. This

chapter examines the bodies of law that are most relevant to these complex transactions and new organizational forms. In considers questions such as: When is an insurer or clinic engaged in the unlicensed corporate practice of medicine? Should physicians who receive capitation payments be regulated as de facto insurance companies? Should HMOs or joint ventures between doctors and hospitals be exempt from tax? What must nonprofit hospitals and insurers do in order to convert to for-profit status? Can hospitals reward physicians for generating income?

A. THE CORPORATE PRACTICE OF MEDICINE

Perhaps the most threatening constraint on organizational innovation in health care is the corporate practice of medicine doctrine, which flatly declares that it is illegal for corporations to pay physicians for medical services. This peculiar doctrine can have devastating legal effects; it can result in the refusal to enforce contracts for medical services and it can lead to injunctions or even criminal sanctions against illegal business arrangements. It is also fundamental to the structure of the American health care delivery system in that it explains why physicians traditionally are financially independent from hospitals, each billing separately for their own services. This division reverberates throughout the system, reflected for instance, in the distinction between Blue Cross and Blue Shield and between Medicare Part A and Part B.

1. THE DOCTRINE'S RATIONALE

The rationale for the corporate practice doctrine is a confusing mixture of formalistic statutory reasoning and policy-based common law. The doctrine's statutory foundation rests on the medical practice act—the physician licensing statute in each state

that makes it a criminal offense for anyone without a license to practice medicine. The doctrine reasons that when a corporation receives money from patients for a physician's medical services, the corporation is engaged in the practice of medicine. The doctrine then observes that corporations, not being natural persons, are ineligible for a medical license because they fail to meet the statutory prerequisites of sound moral character, a medical degree, and a passing score on the medical exam.

The corporate practice doctrine's common law foundation rests on the apprehension that physician employment will lead to debasement of the profession. Courts are concerned that employed physicians will focus unduly on earning a profit, that their patient loyalty will be subverted by their obligation to the corporation, and that their medical judgment will be countermanded by lay owners or administrators.

These rationales have been subjected to sustained criticism throughout its history. The statutory argument is formalistic sophistry in the view of many commentators. The argument is no more sound than the argument, say, that corporations who hire truck drivers are engaged in driving without a license. Clearly, if the *actions* of licensed physicians are to be

attributed to the corporation, their licensed *status* should be as well.

The policy arguments, although more formidable, also are subject to critique. Physicians' financial dependence on corporations may present some marginally increased risk of profiteering and subverted patient loyalty but actual abuses of this nature are not well documented. The corporate practice doctrine arose at a time of widespread quackery in medicine practice, and the doctrine found

its most convincing application in obviously shady cases, such as the corporate dentist who changed his name from Edgar to "Painless." Parker v. Board of Dental Examiners, 216 Cal. 285, 14 P.2d 67 (1932). Nevertheless, courts apply the doctrine even in the most upright circumstances and even when no lay owners or managers are involved. *See, e.g.,* Bartron v. Codington County, 2 N.W.2d 337 (S.D.1942) (illegal to operate a corporate clinic owned by physicians that provided services to indigent patients under contract with the county). As for policy concerns over controlling the course of treatment, admittedly it would be wrong for lay corporate managers to dictate the details of treatment, but such action would itself constitute a direct violation of the medical practice act regardless of who patients pay for medical services. Some degree of corporate influence is tolerable, however, and indeed is encouraged by liability law, which holds hospitals and HMOs responsible for the mistakes of their physicians. It seems contradictory to prohibit the very type of corporate influence that liability law seeks to encourage.

If there is a need for a separate corporate practice doctrine—other than simply to hold a corporation responsible for the acts of unlicensed practice actually committed by its *lay* employees—it is to prevent the possibility of physicians being influenced by corporate profit-making goals. This is viewed by many as naive idealism and out of step with modern public policy which sees a need for economic restraints on medical expenditures. Therefore, the remnants of this doctrine may stand in the path of innovative reform measures necessary to control health care spending. In the view of others, however, the doctrine's policy is far from anachronistic. Instead, the doctrine has renewed purpose in the

modern environment when corporate and economic influences threaten to distort, if not overrun, medical judgment.

Regardless of one's views of the merits, courts have remained virtually steadfast in their recognition of the doctrine. Only two states have rejected it outright. Nevertheless, the extent of the doctrine's presence and enforcement varies widely from state to state. This is due in part to a number of exceptions that have been recognized on judicial and attorneys general opinions. It is also due to differing attitudes and practices among health care lawyers in different states.

2. THE DOCTRINE'S SURVIVAL

The absolutist nature of the corporate practice prohibition seems incredible given the many forms of corporate practice that pervade conventional medical

establishments. Despite the doctrine's unqualified prohibition, physician employment is accepted and common practice in hospitals, government institutions, HMOs, and company clinics.

Explicit statutory exceptions exist for HMOs by virtue of their licensing statutes and the federal HMO Act (42 U.S.C. § 300e–10(a)). In addition, virtually every state has long had a professional corporations law, which allows doctors, lawyers, and other professionals to enjoy the tax benefits of a corporate structure while operating in substance as a partnership. The modern versions of these laws cover limited liability corporations (LLCs) formed by licensed professionals.

Judicially-created exceptions exist in some, but not all, jurisdictions for non-profit institutions and for independent

contracting in contrast with employment of physicians. Potentially the most important exception is for hospitals and other facilities that are licensed under a separate statutory scheme. Some courts, but not all, reason that this suffices to allow corporate payment of physicians within the scope of the facility's license.

Because most of the corporate practice case law dates from the 1930s, there has been some speculation in modern times that the doctrine may have died a quiet death during ensuing decades. However, corporate practice precedents survive in attorney general opinions and in established case law as "legal landmines, remnants of an old and nearly forgotten war, half-buried on a field fast being built up with new forms of health care organizations.

Occasionally, usually at the instigation of those who resist the changes now taking place, one is detonated, with distressing results." Arnold Rosoff, *The Business of Medicine: Problems with the Corporate Practice Doctrine*, 17 Cumb. L. Rev. 485, 499 (1987). Instances of modern application include Carter-Shields v. Alton Health Institute, 777 N.E.2d 948 (Ill. 2002) (nonprofit corporation may not own physician office practices); Conrad v. Medical Bd. of California, 55 Cal.Rptr.2d 901 (App. 1996) (hospital district may not employ physicians).

B. INSURANCE AND HMO REGULATION

1. STATE SOLVENCY LAWS

In most states, insurance commissioners regulate the financial structure of insurance companies and the behavior of insurance sales forces, and in some states they also control the actual pricing and terms of insurance policies. Early on, litigation ensued over

whether precursors to modern-day HMOs were subject to this regulatory authority. Courts initially held no, reasoning that HMOs bore less financial risk since they undertook the limited service obligation to deliver or arrange for care rather than the potentially unlimited financial obligation of indemnifying patients for the costs of whatever doctor or hospital they chose. Jordan v. Group Health Ass'n, 107 F.2d 239 (D.C. Cir. 1939). Nevertheless, these and other laws, including the corporate practice prohibition, were seen as stumbling blocks to HMO development. For instance, many states had licensing statutes for Blue Cross that essentially required non-indemnity

forms of insurance to have the support of the local medical society, and organized medicine has always opposed HMOs with all means at their disposal.

Most of the legal stumbling blocks were removed in a tide of legislation in the early 1970s embracing the HMO concept. The federal HMO Act 42 U.S.C. § 300e created grants and incentives for HMOs that voluntarily meet federal requirements, and preempted most of the obstructive state laws. States adopted HMO enabling statutes that regulate HMOs both as insurance companies and as health care delivery organizations. As a consequence, modern courts find that HMO-type arrangements do constitute the business of insurance.

The overt regulation of HMOs creates a new regulatory dispute, namely, whether risk-bearing arrangements created by provider groups rather than insurance companies are subject to insurance regulation. Hospitals and physician groups are undertaking pre-paid capitation or other risk-bearing contracts with both insurers

(so-called "downstream capitation") and with employers ("direct contracting"). The first arrangement is designed to shift some or all of the insurance risk to the providers. The second is designed to cut out the insurer "middle-man" and to cater to self-insured employers. Under either arrangement, insurance regulators and licensed insurers contend that some providers undertake excessive financial risk for which they should have to meet solvency standards. Otherwise, there is an uneven competitive playing field which subjects licensed entities to greater costs

and oversight and threatens disruption of service in the event of bankruptcy. Provider groups respond that their risk is less than that of insurers, for reasons similar to those advanced decades ago in *Jordan* supra.

The legal response varies and is still evolving. Some states regulate only direct capitation contracts with employers and not sub-capitation contracts that are downstream from insurers, reasoning that in the latter situation there is already regulatory protection for the ultimate consumer. However, ERISA preemption, discussed below, raises some doubt about whether capitation arrangements with self-funded employers are subject to state authority.

2. MANAGED CARE REGULATION

There are two visions of how managed care might improve health care. The *competitive vision* sees in HMOs and similar arrangements the potential for translating cost savings into lowered premiums. The *access vision* sees in managed care the potential for translating savings into increased services. The federal HMO Act initially followed the access model, by requiring as a condition for

receiving federal grants and loans that HMOs engage in community rating rather than experience rating and that they provide comprehensive services with only nominal copayments from consumers. Thus, the main attraction for HMO enrollment has not been lower premiums for the employer but free check-ups, discounted prescriptions, and "first-dollar" coverage

for the patient, all measures that tend to increase rather than restrain utilization.

Managed care insurers now, however, are much more vigorously cost competitive than in the 1980s. The success in containing costs led, though, to concerns about the quality of care. While traditional fee-for-service reimbursement creates incentives for excessive *over*utilization, capitation can create incentives for excessive *under*utilization. Managed care also restricts patients' choice of physicians and physicians' choice of treatment. Nevertheless, most broad-based studies conclude that HMOs deliver care of equal or greater quality than under traditional fee-for-service insurance, and that HMO subscribers are generally well satisfied with the care they receive. There are isolated instances of unscrupulous HMOs that have committed serious abuses, but most of these instances have occurred with HMOs participating in Medicaid, where patient populations tend to be relatively captive and less able to complain or switch providers.

These developments prompted a regulatory "backlash" against managed care, with states enacting various "patients' rights" regulations in the 1990s, often at the behest of provider groups. This generated controversy over whether displacement of market forces and industry self-regulation was truly more in patients'

interests or providers' interests, or both. Consider each of the following examples as either possible confirmation or refutation of these characterizations: (1) requiring a fair and speedy internal appeals process and notice of appeal rights;

(2) requiring those who conduct utilization review to have a medical license in the state in which the treatment is rendered; (3) prohibiting so-called "drive-through deliveries" in which HMOs require mothers to leave the hospital within 24 hours of an uncomplicated child birth; (4) "any willing provider" laws that require HMOs to accept any physician who agrees to the plan's payment terms and meets its credentialing requirements; (5) requiring HMOs to engage in provider credentialing similar to that conducted by hospitals and giving physicians the right to challenge any adverse decisions; and (6) banning so-called "gag clauses" in managed care contracts with physicians, which prohibit physicians from making disparaging comments about the plan or from revealing trade secrets about payment terms or utilization review criteria. Several of these provisions have since been adopted by federal law.

3. FEDERAL INSURANCE REFORM

The Patient Protection and Affordable Care Act of 2010 (Affordable Care Act or ACA) comprehensively reformed how health insurance is designed and sold. This law:

- Prohibits insurers from declining or limiting coverage or setting premiums based on health status or risk, except that older people and smokers can be charged more

- Taxes employers with more than 50 workers that fail to provide affordable insurance benefits

- Provides sliding-scale subsidies to people who do not qualify for Medicaid and do not have affordable coverage through their jobs
- Creates insurance "exchanges" in each state to distribute subsidies and to offer private insurance to individuals and small employers through a mechanism that facilitates comparison shopping
- Determines what "essential" benefits all insurers must cover

Although the Trump administration failed to repeal the ACA, it implemented a number of regulatory changes that critics contend are meant to weaken the law, mainly by expanding opportunities to sell non-compliant policies at lower prices to healthier subscribers.

C. ERISA PREEMPTION

The Employee Retirement Income Security Act of 1974 (ERISA), 29 U.S.C. § 1144, is a federal statute that primarily regulates employer-sponsored pension plans to make sure that employers keep their promises and that funds are fairly administered. Incidentally, ERISA also covers other fringe benefits such as health insurance, but imposes very little regulation on them. In order to prevent inconsistent and overlapping state regulation, ERISA broadly preempts any state law that "relates to" an employer-sponsored fringe benefit plan.

We discuss ERISA preemption at this point because it incorporates the concept of the "business

of insurance" discussed above and it limits the extent of state regulation of health insurance to a significant extent. However, it cannot be overly stressed that ERISA preemption permeates the landscape of health care law and public policy. It has major impact in at least three distinct places: (1) malpractice actions against health insurers; (2) contract claims for the denial of payment under health insurance; and (3) state regulation of how health insurance is structured and sold. The focus here is primarily on the third component, but this section also provides an overview that is useful for the other two topics. Moreover, it cannot be overemphasized that ERISA preemption is capable of cropping up almost anywhere in health law. For instance, each of the following are potentially preempted: physicians' contract actions against managed care plans that drop them from their networks; taxation of firms that assist self-insured employers in administering their health benefits; and state laws that limit health insurers' or employers' subrogation rights when employees' tort awards include medical expenses. *See generally* S. Law & B. Ensminger, Negotiating Physicians' Fees: Individual Patients or Society?, 61 N.Y.U. L. Rev. 1, 80–81 (1986) ("in this judicially constructed Alice in Wonderland world, any state seeking to regulate insurers' arrangements with physicians or providers must be prepared to litigate claims of ERISA preemption").

ERISA preemption is defined by the interaction of three distinct statutory phrases: (1) preemption applies to any state law that "relates to" employee benefits; (2) state authority is restored by an

"insurance savings clause" that allows state regulation of the business of insurance; but (3) under the "deemer clause," states may not deem employers who self-fund rather than purchase insurance benefits to be engaged in the business of insurance. We will look at each of these phrases in turn.

The Supreme Court has stressed that the basic preemption provision, which reaches any state law that "relates to" an employee benefit plan, is to be construed very broadly. Prohibited state laws include both common law and statutory law. Thus, the Court has ruled that employer-sponsored HMOs cannot be sued under state statutory or common law for personal injury damages arising from their negligent limitation of medical care. Aetna Health Inc. v. Davila, 542 U.S. 200 (2004). ERISA itself provides direct remedies against insurers for wrongful limitation of benefits, but ERISA's own remedies are limited to only injunctive and ordinary contract damages (that is, the value of the promised benefits), excluding any punitive damages or compensation for personal injury or pain and suffering. Mertens v. Hewitt Associates, 508 U.S. 248 (1993). That makes sense for a statute designed to enforce pension benefits, but not for the breach of health insurance, which results in the denial of medical care. However, patients still may bring garden variety malpractice suits against HMO *physicians*, or *vicarious* liability theories against HMOs. Generally speaking, these are not preempted because they implicate the employee benefit (health insurance) only indirectly.

237

One notable decision that finds a health care law not potentially preempted is New York State Conference of Blue Cross & Blue Shield Plans v. Travelers Ins. Co., 514 U.S. 645 (1995). In *Travelers*, the Court held that ERISA does not preempt a New York

statute that required hospitals to add a surcharge to their rates in order to fund a pool that reimbursed hospitals for the costs of treating patients without insurance. The plaintiffs in *Travelers* argued that the New York statute was preempted by ERISA because it made health insurance more expensive for employers to purchase, but the Supreme Court disagreed, holding that statutes that have "only an indirect economic effect on the relative costs of various health insurance packages." The Court emphasized, though, that it still supports a broad reading of "relates to."

The sweeping preemption provision is significantly limited by an "insurance savings clause" that reinstates state authority over matters that constitute the business of insurance. The Supreme Court has ruled that the same definition of insurance applies here as under the McCarran-Ferguson Act, which limits the application of other federal laws such as antitrust, securities, and banking regulation to insurance, so that these do not infringe on the traditional role of the states in regulating insurance. Thus, the definition of insurance becomes critical to determining the scope of state authority to regulate health care. For instance, in Metropolitan Life Insurance Co. v. Massachusetts, 471 U.S. 724 (1985), the Supreme Court held that a state law mandating the inclusion of mental health benefits in group

health insurance falls within the savings clause and therefore is not preempted. In so holding, the Court established a three-part test, taken from McCarran-Ferguson Act precedents, for determining whether a given activity constitutes the business of insurance: "First, whether the practice has the effect of transferring or spreading a policyholder's risk; second, whether the practice is an integral part of the policy relationship between the insurer and the

insured; and third, whether the practice is limited to entities within the insurance industry."

These three independent conditions severely limit the scope of the insurance savings clause and thus greatly expand the scope of ERISA's preemption provision. State law may regulate only those activities that fall directly within the core risk-spreading activity of insurance. This limitation excludes a broad array of important activity within the insurance industry. For example, the Supreme Court has held that an insurance company's determination of how much it is willing to pay a health care provider does not fall within the business of insurance because the decision does not relate directly to the policy holder. Group Life & Health Ins. Co. v. Royal Drug, 440 U.S. 205 (1979). However, the Court subsequently ruled that insurers' decisions about which providers to cover at all are part of the business of insurance, in a more common-sense understanding, and thus are subject to state regulation. Kentucky Ass'n of Health Plans, Inc. v. Miller, 538 U.S. 329 (2003). This signals a more lenient approach to preemption.

239

But wait, there's more! A third phrase in the statute declares that states may not deem a benefit plan itself to be insurance. As a result, another important policy effect of ERISA preemption is to promote employer self-insurance. The upshot of this "deemer clause" is that employers who self-fund medical benefits rather than purchase insurance for their employees are entirely exempt from state regulation, regardless of the scope of the savings clause. This provides a strong impetus for larger employers to self-insure their health care benefits, in order to reduce state regulation and avoid the cost of premium taxes. For instance, in *Metropolitan Life* supra, the Court observed that the state could not mandate that self-

funded employers provide the same mental health benefits that are required of regulated insurers. Consequently, the number of employees covered by self-insured health care plans grew dramatically in the 1980s.

To recapitulate: (1) ERISA's preemption clause is extremely broad because so many state laws "relate to" employee benefits. (2) The insurance savings clause would reinstate most relevant state regulation, except for the fact that it has sometimes been given a narrow construction. And, (3) regardless, for reasons just explained, self-insured health benefits may never be subjected to state regulation. The confusion surrounding this complicated scheme and the meaning of its various parts have deterred states from asserting more aggressive regulatory jurisdiction, even where that might be possible to do.

This would not be so troubling if it were not for the fact that substantive ERISA law imposes very little regulatory oversight to take the place of preempted state law. ERISA preemption makes perfect sense for private pension plans, which are closely monitored by the substantive provisions in ERISA, but other employee benefits were added to ERISA as a legislative after-thought, and so ERISA is not designed to regulate or enforce them. For these benefits, ERISA once created a severe regulatory vacuum by failing to fill the void it creates. For instance, McGann v. H & H Music Co., 946 F.2d 401 (5th Cir. 1991) held that neither federal or state law can prevent an employer from switching to self-funded insurance in order to virtually eliminate its coverage of treatment for AIDS, shortly after it discovered an employee was afflicted. The court observed that ERISA was not intended to guarantee any particular level of benefits but only to enforce the benefits

promised, and here the employer had not promised never to change its insurance benefits.

Under the Affordable Care Act, ERISA pre-emption remains intact, but new federal requirements partially fill ERISA's regulatory vacuum, either by regulating some of what self-funded insurers do, or creating incentives for them to conform to federal norms. For instance, federal law still does not require self-funded employers to offer any particular benefits, but it taxes them if they do not offer benefits that meet federal standards for "minimum essential" coverage.

D. CHARITABLE TAX EXEMPTION

Most hospitals are organized as nonprofit corporations. Standing alone, nonprofit status confers no special tax advantage; it means merely that no capital stock is issued and no dividends are paid. However, many nonprofit corporations, including hospitals, are eligible for various forms of both federal and state tax relief because they also qualify for classification as charitable organizations. Principally, section 501(c)(3) of the federal tax code provides that "corporations . . . organized and operated exclusively for religious, charitable, scientific . . . or educational purposes, . . . no part of the earnings of which inures to the benefit of any private . . . individual," shall be exempt from income tax. Nonprofit hospitals are also uniformly exempt from local property taxes and state income taxes. Because nonprofit institutions historically have dominated the health care sector, the requirements for maintaining tax exemption have a profound effect on the organization and operation of health care facilities. The following discussion explores three aspects of tax exemption: the basic eligibility for

charitable status, the prohibition against private inurement, and the effect of receiving income that is unrelated to the institution's charitable purpose. At the end, this section also considers the conversion of no-profit facilities to for-profit status.

242

1. THE BASIS FOR TAX EXEMPTION

a. Hospital Services

Resolving the basic question of why hospitals are considered charitable organizations is more difficult than might first be expected. Hospitals originated during the 18th and 19th centuries as almshouses for the poor, essentially warehousing the impoverished sick who, prior to modern medicine, were not fortunate enough to afford treatment at home. These desperate institutions were usually operated under the auspices of a religious order and were supported almost entirely by charitable donations of time, money and property.

Consistent with these historical origins, the IRS initially required charitable hospitals to provide free care "to the extent of their financial ability." Rev.Rul. 56–185. However, as insurance became widespread for the middle class in the post-World War II era and as the social welfare programs of the 1960s extended broad financial support to the elderly and poor, the optimistic attitude arose that there would soon be no demand for charity care and no need for charitable donations.

How, then, were hospitals to justify continued charitable status? The IRS responded in 1969 by modifying its position to allow hospitals to receive tax-exempt status merely on the conditions that they not discriminate among paying patients and they treat all

indigent *emergency* patients for free. Rev.Rul. 69–545. The IRS went a step further in 1983, ruling that even free emergency care is not

required if an emergency room is not needed or appropriate at a hospital (for example, because it is a highly specialized facility such as an eye and ear hospital). Rev.Rul. 83–187.

Under these rulings hospitals no longer have to provide any specific amount of health care free of charge to the poor in order to qualify under § 501(c)(3) of the Internal Revenue Code. This result is not entirely unprecedented. To be considered "charitable" under the tax code an organization need not necessarily be a charity in the everyday sense of assisting the poor. Many educational and scientific organizations exempt from tax and supported by tax deductible contributions provide no or few services to the poor; serving the poor has never been the only activity accepted as "charitable" under either the tax code or the traditional law of charitable trusts. The IRS consulted precedents from the law of charitable trusts to determine that advancing health provides a separate and sufficient basis for favorable tax treatment, apart from charity care. In essence, hospital care is treated like education—as being a *per se* charitable service.

This shift in policy was opposed by welfare advocacy groups who feared that the naive optimism of the 1960s, which thought the demand for charity care would be fully met by Medicare and Medicaid, was wrong. However, legal challenges to the IRS failed in 1976 when the Supreme Court ruled that patients do not have standing to challenge a hospital's tax status. Simon v. Eastern Kentucky Welfare Rights Org., 426 U.S. 26 (1976). As a result,

many hospitals faced with other pressing financial problems have begun to abandon or diminish their traditional commitment to treating the poor.

A similar battle has occurred at the state level, where advocacy groups have found allies in local taxing authorities hoping to secure new sources of property tax revenue. While state courts have usually followed the federal lead, several notable cases take a sharply different tack, revoking property tax exemption on finding that hospitals fail to provide sufficient services to the poor and to attract a significant level of donations. Also, several states have passed legislation requiring that nonprofit hospitals provide substantial amounts of free services on a charitable basis. Federal law does not set a quantitative requirement, but now requires tax-exempt hospitals to clarify and publicize their charity care standards.

b. Other Health Care Institutions

Hospitals are not the only health care institutions concerned with tax exempt status. Nonprofit health maintenance organizations (HMOs), outpatient clinics, and pharmacies have also sought to maintain exempt status. When confronted with these nonhospital health care services, federal and state tax policy tends to take a distinctly less welcoming attitude.

Despite the IRS acceptance of health care as a *per se* charitable enterprise, it has been reluctant to confer tax exempt status on nonprofit physician groups, HMOs, or other medical enterprises. For

instance, the tax court has ruled that a nonprofit pharmacy organized to sell discount prescription drugs to the elderly does not pursue a charitable purpose. Such an enterprise clearly serves the community's health care needs; nevertheless, the tax court ruled that an exemption was not warranted because the sale of drugs at cost is a "substantial commercial activity . . . in competition with profit-making drugstores." Federation Pharmacy Services, Inc., 72 T.C. 687 (1979). This commerciality standard obviously cannot be reconciled with the *per se* theory that is applied to hospitals, since they too are in direct competition with for-profit firms.

The IRS has also grappled with whether nonprofit health insurers and HMOs qualify for exemption. In 1986, Congress withdrew exemption from Blue Cross/Blue Shield because it viewed the sale of insurance as an inherently commercial activity. I.R.C. § 501(m). How should HMOs then be treated? At first, the IRS took the position that HMOs are merely another form of insurance and therefore not entitled to exemption. However, under pressure from contrary tax court rulings, the IRS adopted a position that seeks to classify which HMOs are more like hospitals, and which are more like insurers. Roughly speaking, the current IRS position is that staff model HMOs are more like hospitals because they directly deliver health care services, and so they can qualify for exemption if they are sufficiently open to the public at large and provide some measure of free services, especially if they own and operate nonprofit hospitals with open emergency rooms. However, IPA or network model HMOs are not exempt, even if they

meet these same standards, because they merely "arrange for" health care services. In the leading case of Geisinger Health Plan v.

Commissioner, 985 F.2d 1210 (3d Cir. 1993), the Third Circuit sustained this position.

A related issue of dispute has been whether integrated delivery systems (IDSs) can receive charitable exemption. They are composed of large networks of physicians and hospitals and they sell insurance, so they might fit within any of several different precedents. For the most part, the IRS has been accommodating to IDSs, applying to them essentially the same tests as apply to hospitals, especially if the network is built around a large nonprofit hospital that also supports education and research functions. However, in one respect, the IRS rulings have made exemption difficult for integrated systems. The IRS initially ruled that exempt status would be questioned if more than 20 percent of the governing board is composed of physicians who practice in the system. This ruling was based on obscure precedents from tax exempt bond financing, and is premised on the idea that physicians have a conflict of interest that might potentially cause them to manage the system for their own professional benefit rather than for the community's benefit. The consequence, however, is that many integrated systems opted against nonprofit status because they felt they could not operate and recruit effectively unless physicians were given a significant leadership role. Under harsh criticism, the IRS relented on the 20 percent guideline and now allows up to 49 percent physician membership on the board, as long as

physicians' salary decisions are made by an independent committee and there are other conflict-of-interest protections in place.

The sharp dichotomy between exemption for hospitals and exemption for other health care services reflects the reality that neither state nor federal taxing authorities actually accept the proposition that health care is a per se charitable purpose, just as they would be unlikely to exempt a nonprofit bookstore despite the per se exemption for "education" that is explicit in the statute. Instead, these authorities are searching for an alternative exemption rationale to differentiate between those health care services that deserve a tax subsidy and those that do not. Although the basis of individual rulings may (or may not) be fairly clear, the overarching theory of charitable exemption remains elusive.

2. INUREMENT TO PRIVATE BENEFIT

In addition to serving a charitable purpose, health care institutions must comply with several operational requirements to be eligible for tax exemption. The principal threat to exempt status is the private inurement prohibition: "*no* part of the [hospital's] net earnings [may] inure[ ] to the private benefit of *any* private shareholder or individual." Code § 501(c) (emphasis added). For example, in the leading federal decision, the court found that restricting hospital access exclusively to a small group of physicians undermines exempt status because it converts the hospital into essentially a

private workshop rather than a public facility. Harding Hospital, Inc. v. United States, 505 F.2d 1068 (6th Cir.1974). The court also cited as evidence of private inurement: (1) the hospital's provision of office space, medical equipment, and clerical services to staff physicians at below-market rates; and (2) the hospital's payment of a large management fee to the doctors for hospital supervision.

The IRS has identified several other activities that may jeopardize exempt status. Hospitals frequently offer a package of benefits to new physicians in order to attract them to understaffed rural areas, benefits such as discounted office space and a guaranteed minimum income. The IRS has indicated that such inducements may or may not constitute private inurement, depending on the reasonableness and necessity of the amounts offered in individual cases. Similarly, the IRS has stated that it may be legitimate for a hospital to act as a general partner with physicians in a venture to construct a medical facility, even though by so doing the hospital contributes substantial capital and places its assets at risk, but such joint venture activity requires a highly individualized determination of the value of the *quid pro quo* in each case.

Where the IRS has drawn the line is at joint ventures designed solely to give physicians a financial stake in the success of an existing department or facility. This can occur through arrangements that blatantly sell to physicians for a nominal amount a portion of the revenue stream produced by their hospitalized patients. In General

Counsel Memorandum 39862 (Dec. 2, 1991), the IRS ruled that this constitutes obvious private inurement because there is no shared investment risk in a new facility or service, only a thinly disguised attempt to encourage physicians to make greater use of one hospital rather than another. This is a motive the IRS does not consider to be a benefit to the community, even though it clearly does benefit the hospital.

3. UNRELATED BUSINESS INCOME

A matter of less serious concern, but still of substantial importance to tax-exempt organizations, is the treatment of income from activities unrelated to the entity's exempt purpose. Such unrelated business income does not jeopardize the organizations' *overall* exempt status unless it rises to a substantial level in comparison with overall operations. Nevertheless, unrelated business income is subject to separate taxation. For instance, income from a hospital gift shop would never be large enough to threaten taxation of the hospital's entire income, but if gift shop income were considered "unrelated" to a hospital's exempt purpose, the gift shop income itself would be taxed. The sources of such potentially unrelated income in hospitals are numerous and significant enough to have generated a number of rulings.

The Code provision on unrelated business income, section 513(a), defines such income as that "which is not substantially related [to the] . . . purpose or function constituting the basis for exemption, . . .

except that such term does not include any trade or business which is carried on . . . by the organization primarily for the convenience of its . . . patients or employees." Consider how this definition might apply to income from hospital gift shops, cafeteria sales, and parking lot receipts—services that are patronized largely by visitors. The IRS has found these sources of income to be exempt. This result is easily justified under the "convenience for patients and employees" exception, but the Service has chosen to rest this result in the strained reasoning that such sales are substantially related to patient care by virtue of the therapeutic benefits of patient visitation. The IRS takes the opposite stance when it addresses income from hospital pharmacies and laboratories

derived from nonhospital patients. The Service has ruled that this income is *un*related to a hospital's exempt purpose, even though it is patient care income from patients of doctors on the medical staff whose offices are maintained next door in a tax-exempt office complex.

4. HOSPITAL REORGANIZATION, DIVERSIFICATION, AND CONVERSION OF STATUS

a. Reorganization and Diversification

The tumultuous environment that prevails in the health care industry is prompting widespread experimentation with organizational innovations. Virtually every hospital of any significant size has undergone some type of corporate reorganization in the past generation. Typically, a nonprofit hospital

will segregate its various functions, some for-profit and some nonprofit, and incorporate them into separate entities, all under the common ownership of a holding company. Such corporate restructuring raises many of the tax issues discussed so far, such as whether a particular hospital function (say, laundry) standing alone will support an exemption and, if not, whether it is sufficiently related to an exempt function to be free from the unrelated business income tax if organized as part of the hospital corporate entity.

The tax consequences of hospital diversification are especially important in the modern business climate. The rampant excess hospital capacity and shrinking patient service revenues have caused hospitals to venture into all sorts of new enterprises. Some are directly related to health care, such as nursing homes and home health services. Others are far afield, such as athletic clubs, real

estate development, and day care centers. Sophisticated hospital counsel, in deciding how to structure these ventures, must not only consider the tax problems surveyed here but also the potentially countervailing effects of government reimbursement regulations and certificate of need requirements. Tax law might prompt a hospital to internalize a new venture in order to claim it as related business income, but doing so might necessitate a certificate of need or might adversely affect reimbursement under Medicare, Medicaid, or state rate regulatory programs (topics addressed elsewhere in this book).

b. Conversion to For-Profit Status

Another type of restructuring that has swept the industry is conversion to for-profit status. Despite the benefits of tax exemption, many hospitals, HMOs, and Blue Cross plans are finding it more attractive to forgo nonprofit status. Perhaps this is prompted by the demands of maintaining charitable tax exemption, or perhaps it is that for-profit status allows better access to capital markets by permitting equity investment. Another possibility is that members of nonprofit hospital and Blue Cross boards, who often view their volunteer positions as community service, are finding the modern climate too difficult to continue operations with a truly charitable mission. For whatever reason, legions of nonprofit conversions have occurred, generating enormous public controversy. The same is true for public hospitals, many of which have converted to private status, often to avoid restrictive enabling legislation that keeps them from entering into the contractual arrangements that are necessary to compete in a managed care environment.

Conversion of status can occur through several different forms. The simplest is for the nonprofit corporation or municipal government to sell its principal asset to a for-profit entity so that now it owns cash rather than a hospital or insurance company. Somewhat more complex is to retain ownership of the facility but enter into a long-term management contract with a for-profit firm to run the hospital, rewarding it based on a percentage of the profits. Either type of transaction raises a host of

corporate or municipal law issues that must be resolved.

The first issue is whether the transaction is within the corporate powers or instead is *ultra vires*. This issue arises because the articles of incorporation for nonprofit entities, and the authorizing legislation for government entities, frequently limit the purposes for which the entity can operate. For instance, if these governing documents authorize hospital operations but not health care or charitable purposes more generally, then the entity has difficulty justifying selling a hospital and devoting the proceeds to some other use. Similarly, if donors to the hospital have restricted their gifts to particular uses, such as supporting a particular facility, then any conversion of the assets would violate a promise made to donors who may be long-since deceased.

One way nonprofits can avoid the restrictions of the *ultra vires* doctrine is to invoke the concept of *cy pres* from the law of charitable trusts. The *cy pres* doctrine addresses situations where it is no longer possible or feasible to carry out the original terms of a charitable gift, and so courts can authorize a change in the purpose to the next closest use. Whether the intensely competitive climate of modern health care is a compelling enough change in

circumstances to force a charitable hospital to abandon its nonprofit status is a question on which few courts have ruled, but modern trust law decisions generally apply the *cy pres* doctrine somewhat more liberally than in the past. However authorized, once charitable assets are sold, then these same rules

control how the proceeds can be used. Most nonprofit hospitals, HMOs, and Blue Cross plans have chosen to use the proceeds to fund a charitable foundation to advance community health.

The next issue, then, is whether the amounts received in the transaction are sufficient. This refers to the fiduciary duties of care and loyalty that all corporate managers, including nonprofit board members, owe in their corporate governance. The risk that exists in nonprofit settings is that the absence of any oversight from owners and shareholders with a stake in the transaction may make board members lax in their negotiations. Even more troubling, the acquiring for-profit firm may undermine nonprofit board members' loyalty by holding out the prospect of rich personal rewards if the transaction goes forward. For instance, board members have been offered lucrative positions or bargain basement ownership shares in the new for-profit entity, which obviously encourages them to approve the transaction. Fiduciary law that applies to nonprofit and for-profit corporations alike prohibits such conflicts of interest and can be used to void tainted transactions.

Resolving these issues requires that nonprofit assets be fairly valued in conversion transactions. Fair valuation of nonprofit assets is a difficult analytical exercise, however, on which opinions differ widely. One possibility is a "hard asset, depreciated value" approach, which looks at what it would cost to build or replace the

assets. This tends to produce a fairly low valuation because it ignores the firm's good

will and established reputation in the market. A second approach is to value the enterprise as an ongoing business, but even this produces a valuation that is often much lower than the speculative value that can result from the sale of its stock. There are numerous instances of hospitals and insurers whose stock market value a year or two after their conversion ends up being many times higher than the sales price. This creates a strong impression that the original price was greatly depressed due to a lack of arms length or conflict-free negotiations. On the other hand, a good bit of this value can be attributed to the mere fact of the conversion. The enterprise is simply worth more now that it can have equity investors and its management is in different hands. It is not clear who should get credit for this added value that is created by the transaction itself.

The number and difficulty of issues presented by a conversion transaction raises the further question of who is in a position to make sure they are properly resolved. Owing to the lack of shareholder oversight, state law usually gives the attorney general authority to challenge in court any nonprofit transaction that appears questionable. In the past, careful counsel sometimes requested and paid for an attorney general to bring a nominal challenge simply in order to gain the court's approval so the transaction cannot later be invalidated. Increasingly, however, attorneys general have posed real opposition to conversions, reflecting the public's hostility to losing the nonprofit character of valued community institutions. In a number of states, legislatures have required approval by attorneys

general or administrative agencies before conversions can proceed, and attorneys general themselves have been proactive in promulgating guidelines and negotiating the terms under which they will approve, or not oppose, conversions.

E. REFERRAL FEE PROHIBITIONS

1. SOURCES OF LAW

Paying health care providers purely as an inducement to refer patients has long been considered unethical and inappropriate. Kickbacks for ordering or recommending treatment have been shown in various empirical studies to distort judgment about whether services are necessary and who are the better providers. These costs are borne not only by the patients affected, but also by the public and private insurance that pays for the care. Accordingly, referral fees are prohibited by several sources of law.

Fee splitting is frequently listed in medical practice acts as one basis for professional discipline. The payment of referral fees also constitutes a felony under some state statutes. *E.g.,* Cal.Bus. & Prof.Code § 650. But the federal law contains the most prominent prohibition. The anti-fraud and abuse provisions of the Medicare and Medicaid programs, paraphrased, declare: anyone who receives or pays any remuneration directly or indirectly, overtly or covertly, in cash or in kind for the referral of a patient to a person for the furnishing of any item or service for which payment may be made under Medicare or

Medicaid is guilty of a felony punishable by five years imprisonment or $25,000, or both. 42 U.S.C. § 1320a–7b(b). Referral fees are also prohibited by a federal statute known as the "Stark law," named for Rep. Fortney "Pete" Stark who championed it. The Stark law prohibits Medicare and Medicaid from paying for any services that physicians render or order through entities that they have a financial relationship with. 42 U.S.C. § 1395nn. This is meant to prohibit so-called "self-referral fees," that is, incentives for physicians to send patients to entities they own or receive money from.

So, the message is clear: referral fees are bad. Unfortunately, confusion reigns over precisely what this means. Implicit or explicit referral incentives pervade most accepted and legitimate relationships in medicine, so it is far from clear which are prohibited and which are not. For instance, it could be argued that simply offering a discount in the sale of medical services constitutes a sort of self-referral fee since the discount might, quite literally, be characterized as a "kickback" from the seller to the purchaser intended to induce the purchaser to select that seller. As another example of completely innocent referral incentives, a hospital's granting of medical staff privileges might be viewed as a form of consideration meant to encourage physicians to admit patients to the hospital. Yet it would be ludicrous to contend that this cornerstone of the health care delivery system is fundamentally unethical and criminally illegal. It is necessary, then, to search for some analytical guide to distinguish illicit referral incentives from acceptable practices.

2. EARNED VERSUS UNEARNED FEES

The analytical distinction that provides the greatest assistance in determining which referral incentives are illegal is to distinguish between earned and unearned fees. Earned fees are those in exchange for a legitimate, non-referral service of fair market value. Unearned fees are any that do not have a non-referral quid pro quo or that are in excess of the fair market value for the exchange service.

This is a quite sensible and workable approach that avoids messy inquiries into subjective motives. For example, California has ruled that radiologists and pharmacists can be required to pay hospitals a reasonable percentage of the revenues they generate from hospital patients in exchange for the hospital providing equipment and staffing, despite the fact that this results in the physicians paying the hospital in proportion to the amount of business received from hospital patients. If payments are commensurate with the hospital's actual expenses in furnishing facilities and support, they are regarded as fees for services rather than fees for referrals.

Unfortunately, referral fee interpretations frequently stray from this analytical path. The decision that has caused the most concern under federal law is United States v. Greber, 760 F.2d 68 (3rd Cir.1985). It addressed a clinical laboratory's practice of paying "interpretation fees" to compensate physicians for evaluating their patients' test results, a well-recognized disguise for a kickback. The court sustained a criminal conviction, observing that the fee paid exceeded the value of the interpretation

services and that some doctors were paid even though they didn't perform any service. Although this decision was clearly correct on its facts, the court made a broad and troubling pronouncement that

reached far beyond the particulars of the case: "if the payments were intended to induce the physician to use [the laboratory's] services, the statute was violated, *even if the payments were also intended to compensate for professional services*." 760 F.2d at 72 (emphasis added). *Greber* thus leaves in doubt what analytical guide should be used to reach a sensible interpretation of this broad and ambiguous concept.

Additional guidance comes from several different sources. Under the Medicare/Medicaid anti-kickback statute, DHHS is instructed to promulgate safe-harbor regulations "specifying payment practices that shall not be treated as a criminal offense under [the referral fee statute]." These safe-harbor regulations rely heavily on the fair market value concept. For instance, they allow physicians to invest in medical enterprises if they do so on terms that are offered to non-physician investors, and they allow hospitals to rent space to physicians on terms that are "consistent with fair market value in arms-length transactions." Similar guidance is contained in the Stark statute itself, through a number of complex statutory definitions and exceptions that capture the same concepts. However, these regulations and statutory provisions are narrowly drawn to cover only a limited number of discrete transactions, and only those that are generally accepted to be beyond approach. They do not offer more general analytical guidance for resolving innovative arrangements that

are not yet in common use or those that are in use but that are more difficult to judge. This has led to continuing criticism that these laws stifle productive innovation or cause legitimate and well-meaning ventures to operate under a cloud of potential illegality.

Responding to this criticism, Congress has taken two important steps. First, it has required the enforcement agencies to create more sweeping exemptions for incentive arrangements within managed care or capitated payment systems. This recognizes the fact that productivity or inflationary incentives may be an important safeguard to counteract larger cost containment pressures. Second, Congress has required the enforcement agencies to institute a process for obtaining transaction-specific advisory rulings, much like is done under tax and antitrust laws. Although this process is still new, the initial rulings indicate that the enforcement agencies will continue to take a very cautious view of which arrangements are permissible. For instance, in one of the early rulings, a hospital was told that it could not restock supplies on independent ambulances that served its emergency room, for fear that overly generous restocking could be used to encourage ambulance drivers to select one hospital over another.

F. SUMMARY

The array of transactions and organizational structures among doctors, hospitals and insurers is subject to a complex tapestry of legal doctrines. The health care lawyer's job is made more difficult because these multiple bodies of law often give sharply inconsistent signals about which particular arrangements are legally preferred. No one arrangement is ideal, either from a legal or a business planning perspective. To gain some sense of the overall picture that results, consider the following chart, which displays a rough guide for how various business and legal perspectives would regard three different methods a hospital might pursue for closer affiliation with physicians. Suppose the affiliation were undertaken either to market a comprehensive managed care insurance plan, or,

less ambitiously, to establish an outpatient clinic. This chart indicates whether each business or legal factor would view each arrangement favorably (+), negatively (-), or in a neutral/mixed (/) fashion.

262

	Physician Autonomy	Clinical and Financial Integration	Corporate Practice of Medicine	Antitrust	Referral Fee Law	Tax Exempt
Foundation Plan (The hospital's corporate parent buys out physician practices and employs the physicians)	-	+	-	+	+	+
Physician-Hospital Organization (PHO) (Hospital and several physician groups form joint-venture partnership, contribute equal capital, and split the proceeds)	/	/	/	/	-	/
Management Services Organization (MSO) (Hospital contracts with independent physicians to provide office management services and to act as negotiating agent with insurers/employers)	+	-	+	-	/	-

This rough outline is only a general approximation and is subject to some dispute, but for what it is worth, here is what these labels are based on:

- Physician autonomy favors looser forms of integration whereas management concerns favor tighter integration, and so these two dimensions run in opposite directions.
- The *corporate practice doctrine* favors MSOs because payments aren't for medical services. The corporate practice doctrine is most clearly

263

violated by the Foundation model, so that can be used only if state law creates an exception for nonprofits or for hospitals. PHOs may be safe if physicians and hospitals bill and collect separately.

- *Antitrust* favors the Foundation model because it is a single entity, whereas the MSO is the least integrated and therefore potentially closest to price fixing. The PHO would depend on how the particular arrangement falls under the DOJ/FTC guidelines. These considerations don't take account of market share (merger) or boycott (exclusion) issues.
- *Referral fee laws* also favor the Foundation model because payments occur within an integrated entity where it is less obvious that any "referral" occurs, and for which various exceptions exist. PHOs do the worst under the Stark law, although not too badly under the anti-kickback statute as long as the investment arrangement is properly structured. MSOs run the risk of both under-and over-compensating for services, thereby creating potential referral incentives in either direction.
- *Tax exemption* favors hospitals and disfavors doctors. MSOs don't provide medical services and therefore have only a tenuous connection with the health care exemption.

PART III

ETHICAL ISSUES IN PATIENT CARE DECISIONS

In Part III we shift our focus to issues that arise in individual patient care decisions. This Part addresses certain recurring ethical dilemmas that confront physicians and hospital administrators. We begin in Chapter 6 by looking at the problem of defining death, which is related to how organs are harvested and distributed for transplantation since the dead are a critical source of organ donations. A clear grasp of the definition of death is also important to understanding issues that arise in the termination of life-sustaining treatment, which is the subject of Chapter 7. Chapter 8 then shifts our attention to the beginning of life, examining the ethical dilemmas arising in two selected areas of reproductive medicine: advanced reproductive technologies, and maternal-fetal conflict. While Part III provides a comprehensive overview of the relevant law, it often goes beyond the law to survey the philosophical and policy debates that permeate these questions.

CHAPTER 6
DEFINING DEATH AND TRANSPLANTING ORGANS

A. DEFINING DEATH

Death is not necessarily a univocal concept. Medically, death is a diagnosis. Philosophically, the meaning of death is a moral question. Politically, death defines the point at which people's interests, and the obligations owed them, disappear. In religious traditions, death may simply be a transition, a time when the soul leaves the body for another world.

Legally, many different things may turn upon a determination of death: whether a homicide has been committed; when organs can be harvested for transplantation; when burial, cremation, or autopsy can occur; the sequence in which wills should be probated; when a spouse can remarry; whether people have standing to bring a lawsuit in their own right. In theory, one might look to varying criteria for the definition of death, depending on the legal purpose for which the inquiry is made, for example, to determine whether to remove a heart versus whether a homicide has been committed. The law, however, has ordinarily accepted the medical diagnosis as dispositive for all purposes. In the 1970s, a change in the medical view of death was widely and rather quickly adopted by the law. This has prompted an ongoing debate over the proper standard and its implications.

1. CARDIOPULMONARY CRITERIA

Traditionally death was declared when a person stopped breathing, and the heart stopped beating. The shortcomings of this "cardiopulmonary" definition emerged with the advent of ventilators and other technology capable of artificially sustaining one's "vital signs" for days, weeks, or even longer, even though the brain has ceased functioning. Since such persons breathe and their hearts pump blood, the cardiopulmonary standard precluded their being considered "dead" despite their general lack of responsiveness to any stimuli and their inability to sustain even a *vegetative* existence without the technology. In a precursor to the current "futility" debate (see Ch. 7.I), physicians and ethicists questioned the legal standard that treated patients as "alive" simply because some cells or organs continued to live. In people without brain activity, their bodies have permanently lost their capacity to function as integrated wholes. Of at least equal importance, treating such people as alive prevented effective retrieval of their organs for transplantation. While the need to procure organs, by itself, is not sufficient justification for altering the medical definition of death, it provided an important motivation for rethinking how we define death.

2. NEUROLOGICAL CRITERIA: WHOLE BRAIN DEATH

The solution to these difficulties has been to redefine death so that it is properly diagnosed when there has been either a permanent cessation of all

respiratory (i.e., pulmonary) and circulatory (i.e., cardiac) function *or* when the brain has died. The modern legal rule, applied in most states, is stated in the Uniform Determination of Death Act, as

proposed in 1981 by the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research:

> An individual who has sustained either (1) irreversible cessation of circulatory and respiratory function, or (2) irreversible cessation of all functions of the entire brain, including the brain stem, is dead. A determination of death must be made in accordance with accepted medical standards.

All states, by statute or judicial decision, have accepted some form of the whole-brain death criterion.

Note that determining death by neurological criteria requires *whole* brain death, "including the brain stem." Anatomically, the brainstem is an extension of the spinal cord into the base of the skull; functionally, it is responsible for most of the "vegetative" functions essential to maintaining life, such as breathing, swallowing, digesting, and sleep-awake cycles (in contrast with the upper brain, which controls consciousness, thought, feeling, and memory). Thus death of the brainstem means loss of the body's ability to maintain respiration. Without respiration, the heart (which is *not* directly dependent on the brain to continue beating) will die from a lack of oxygen, and circulation will stop. Brain death, therefore, inevitably leads to the termination

of cardiopulmonary functions, unless respiration is artificially maintained.

Because in the normal course of events the three systems (heart, lung and brain) fail more or less simultaneously, regardless of which one fails first, the two formulations are sometimes characterized not as different *concepts* of death, but as alternative

criteria for measuring the same underlying status—the permanent "disintegration" of the organism as a whole, reflected in the cessation of integrated functioning among these three essential, interdependent systems. Some argue, however, that brain death *is* conceptually different, and in a manner that is not simply an alternative to cardiopulmonary death. Due to the brain's unique primacy in integrating the functions of the organism as a whole, its death is the "real" (or at least, the most important) criterion for the demise of a person. Without a living brain, some tissue may be alive, but the *person* is not. In this view respiration and circulation retain political validity only as a concession to lay perceptions of human vitality, and scientific validity only as an indirect measure of brain function.

Whatever the conceptual model, acceptance of brain death is quite important as a practical matter, for it enables the transplantation of organs essential to life. Under current law, taking essential organs from people who are still alive, even with their consent, is homicide, but once the patient's heart stops beating, the donor's organs will soon die from the lack of circulation. While medical advances have

made it more possible to retrieve organs after the cessation of cardiac function (for example, when a patient is moved to an operating room before life-sustaining treatment is withdrawn), organ retrieval is more successful when the patient is dead, but still breathing and circulating blood. The legal acceptance of whole brain death permits this.

Conceivably, one could achieve the same result in other ways, for instance by making an exception to the "dead donor" rule for

terminally ill patients, or by modifying the cardiopulmonary criterion to refer to loss of *spontaneous* respiratory function, but there are problems with each of these approaches. Modifying the dead donor rule would breach the fundamental prohibition of suicide and would pose difficulties in defining terminal status. Modifying the cardiopulmonary standard would pose similar difficulties in determining whether the inability to breathe on one's own is "irreversible;" some people can be "weaned" from ventilators, but it is difficult and harrowing to determine who they are by trial and error. Adopting a brain death standard accomplishes the same result with much greater diagnostic certainty, since a person who has no brain-stem activity also has no spontaneous respiration. Moreover, brain death is morally appealing to those who believe it is conceptually superior to adopt a criterion that focuses on the centrality of the brain in the integration of the organism.

3. NEUROLOGICAL CRITERIA: UPPER BRAIN DEATH

Laypersons sometimes think that brain death occurs simply with the loss of cognitive function (i.e., consciousness). That is clearly wrong under existing law. This point has enormous practical importance. The upper brain, which is the center of cognitive function, is more sensitive than the brainstem to interruptions in the oxygen supply (as may occur in a stroke or a near-drowning). It therefore often happens that the upper brain dies, or suffers permanent damage, while the brainstem remains relatively intact. The person then lapses into a "vegetative state," in which all cognitive function is lost even though the brainstem functions and maintains spontaneous breathing. If physicians believe the upper brain has really died, the patient will be diagnosed as being

permanently unconscious or in a "persistent vegetative state" (PVS). Treatment decisions for patients in such a condition can present difficult ethical and legal problems—indeed, many of the leading end-of-life court cases involved patients in a PVS, including Karen Quinlan, Nancy Cruzan and Terri Schiavo. The issues these cases raised are addressed in the next chapter. However, PVS patients are not dead. (Likewise, patients with an injury to the brain *stem* that is not fatal to that tissue—even though it may compromise spontaneous respiration or other vegetative functions—are not dead.)

Some have nonetheless advocated changing the law to replace whole brain death with *upper* brain

death. An important argument offered for this proposal is that personhood necessarily depends on the higher-brain functions of consciousness, rationality, and the capacity for mental or social functioning. In this view, brain death is really a marker for the true meaning of death, the loss of consciousness.

The President's Commission and other authorities have rejected this position. They point out that a cognition-based standard might result in defining newborns, or the severely mentally handicapped or demented, as dead, or at least not fully human. There is also doubt that we really would be willing to treat permanently unconscious patients as dead. Would we be comfortable burying them without further ado, even while they are breathing? In sum, it is a serious business to label someone dead, and we are properly cautious in extending the definition.

Medical uncertainties would also arise in attempting such a definition. Even if we could determine precisely which portions of

the brain house the cognitive functions essential to being called "alive" or a "person," we would sometimes have difficulty determining with sufficient certainty when, in fact, they have lost function irreversibly. See Chapter 7.G.2 (discussing the report of a woman who woke up from a four-month coma, shortly after a judge ordered her feeding tube removed); The Multi-Society Task Force on PVS, *Medical Aspects of the Persistent Vegetative State,* 330 New Eng. J. Med. 1572 (Second of Two Parts) (1994) (discussing prognosis of those in PVS of varying origins). Such

errors are unavoidable, but we certainly want a definition that minimizes the frequency with which we mistakenly label someone dead.

In fairness, even the diagnosis of whole-brain death is not entirely without difficulty. It is supposed to reflect the complete loss of the brain's ability to exhibit integrative capacity, but brain dead people often maintain some integrative capacity. Their brains can regulate the amount of water retained by the kidneys to keep blood from being either too dilute or too concentrated, and they may exhibit an increase in heart rate and blood pressure from surgical incisions to remove their organs. In addition, the concept of integrative capacity does not explain very well how people think about death. When a pregnant woman dies according to neurological criteria, doctors sometimes maintain her cardiopulmonary function artificially until her fetus is viable and can be delivered by cesarean section. After her child's birth and the discontinuation of the medical treatment, lay people typically will then view the woman as having died. Even physicians and nurses involved in organ transplantation do not think in terms of integrative capacity. When a study asked a group of these physicians and nurses why a person is dead when the brain

stops functioning, only 25 percent cited the loss of integrative capacity, while 36 percent spoke in terms of loss of consciousness, and 32 percent explained that the patient would die soon no matter what was done or that the patient's quality of life was unacceptable. Youngner, *Defining Death*, 49 Arch. Neurol. 570 (1992).

275

Though the law is settled, a vigorous ethical and philosophical debate continues over the proper definition of death. And cases arise in which the family rejects a diagnosis of death, either because they think it was incorrectly made or because they reject the concept of death by brain criteria. In a prominent case, Jahi McMath, at age 13, was declared dead in 2013 in California after complications from a tonsillectomy for sleep apnea. Her family did not believe she was dead, and medical experts disagreed as to whether she was in fact dead. The family moved McMath to New Jersey, whose determination of death statute includes

> The death of an individual shall not be declared upon the basis of neurological criteria . . . when the licensed physician authorized to declare death, has reason to believe . . . that such a declaration would violate the personal religious beliefs of the individual. In these cases, death shall be declared, and the time of death fixed, solely upon the basis of cardio-respiratory criteria. . . . [N.J. Stat. Ann. 26:6A–5.]

About 4½ years later, McMath was declared dead by cardio-respiratory criteria.

In a society characterized by pluralistic religious and philosophical values, some have argued that we could plausibly leave it to individuals and their families to choose and apply their own definitions, perhaps within a range of acceptable

predetermined alternatives (e.g., death in terms of consciousness, brain function, or cardiopulmonary function), or on

the basis of a default rule that individuals would be free to alter within that same range.

B. ORGAN PROCUREMENT AND ALLOCATION

This section explores, first, the basic legal rules governing how organs are obtained, and second, the system for the allocation of transplantable vital organs. The rules and policies governing both questions arise against the background reality that the need for vital organs persistently, and increasingly, exceeds their supply. In 2019, over 113,000 people were on organ transplant waiting lists. Between 2011 and 2018, 6,000 to 7,500 people died each year while on those lists. (For related data, see http://www.unos.org). Even greater numbers of would-be transplantees never make it onto a waiting list (see sec. B.2., *infra*) since, due to the supply shortfall, it is clear from the outset they will not qualify.

This gap arises from many factors. On the supply side, people are hesitant, for many reasons, to fill out donor cards while still alive. After a death, physicians may fail to ask family members for permission to harvest organs, and when they do ask, family members may decline. Even when the deceased person expressed a preference for donation, physicians often defer to any objections from the family. Thus, although formal statements of a desire to donate one's organs upon death are *legally* controlling, Rev. Unif. Anatomical Gift Act § 8(a) (2006), these may be disregarded in favor of family

members' views. On the demand side, while the number of transplants is growing, the waiting list is growing much faster. This may have to do with a simultaneous reduction in the number of medically suitable donors, through factors such as increased highway safety—seat belt laws, air bags, motorcycle helmets, enforcement of drunk driving laws—and an increase in the pool of potential recipients, through medical advances that render more people good candidates for transplantation.

1. PROCUREMENT OF ORGANS

Laws and public policies that seek to procure more transplantable organs must balance the interests of donors, recipients, and the public health. In doing so, they confront a number of important and recurring dilemmas, including the proper role of financial incentives in motivating donations, the definition of death, the proper role of consent, and obligations to donate organs.

a. Organ Donation

Competent living persons may donate renewable tissues (blood, sperm) and those in ample supply (eggs), but not those necessary for sustaining their own life (entire liver, heart, both kidneys). While transplant centers will facilitate donation of a single kidney (or, less commonly, a lung, partial liver, or partial pancreas), particularly between family members, there are surgical as well as medical risks to the donor; life with a single kidney, for example, leaves the donor more vulnerable. There are also

policy criticisms of this practice: within families there is a concern about coercion, and outside families, a concern about "under the table" payments. The latter implicates a federal prohibition of the

sale of any transplantable organ (excluding blood, sperm or eggs), contained in the National Organ Transplant Act of 1984, 42 U.S.C. §§ 273–274g, which supplements traditional state-law regulation of the field.

Postmortem donations are governed by state law versions of the original (1968) or revised (1987 and 2006) Uniform Anatomical Gift Acts. Under these acts, which differ somewhat but have been adopted in some form in all states, competent adults may make gifts of their organs, effective at death, for education, research, or transplantation. If a deceased person neither left instructions to make such a gift nor forbade it, the family may consent to organ harvesting. Starting with the 1987 revised version, hospitals are required not only to make routine inquiry of patients regarding their wish to be postmortem donors, but also to ask families to exercise their authority to permit such donation (where the donor has not acted to prohibit this). The latter provision—"required request"—was designed to remedy physicians' reluctance to raise the question of donation with family members; it is also imposed on hospitals by federal law as a condition of participation in Medicare and Medicaid, 42 U.S.C. § 1320b–8. Organ procurement organizations have been able to raise the consent rate for family members by improving the way they discuss the

possibility of donation. Finally, as under federal law, state law prohibits the sale of organs.

The legal prohibition against selling, during life, a vital organ whose removal is survivable (e.g., single kidney), *or* against contracting for the sale of a heart, lungs, liver, or both kidneys upon death, is justified on several theories. Supporters of the ban

argue that a market approach would "commodify" the body in objectionable ways, undermine altruism, and (in the context of organ removal during life) encourage needy persons to take inappropriate risks for pay. Advocates of a free market approach argue in response that the current system inhibits needed expansion in the supply of organs, and that it is unjust to donors since recipients, physicians, and others all derive concrete benefits from the donors' largesse, yet donors remain uncompensated in material terms. Some proposals would provide payment to family members of people who agree while alive to organ donation after their deaths, to avoid the concern of people putting their health at risk for money.

Observers also have noted that the payments would address real problems with the donative element of organ transplantation. Providing an organ to someone can create complicated, enmeshing relationships—what has been called the "tyranny of the gift"—among the donee, the donor (if alive), and their respective families. Renee C. Fox & Judith Swazey, SPARE PARTS: ORGAN REPLACEMENT IN AMERICAN SOCIETY 40–41 (1992). In recent years, public and professional sentiment has shifted enough

in favor of some financial incentives that states have started to permit income tax deductions for people who donate a kidney, partial liver, bone marrow, or other organ. The tax deductions can be claimed for unreimbursed expenses, including lost wages. In December 2019, the U.S. Department of Health & Human Services proposed a new regulation to expand the scope of reimbursable expenses (e.g., childcare and eldercare expenses). The debate is ongoing, and merely summarized here.

b. Donation and the Definition of Death

More organs would be available for transplantation if the permanently unconscious were defined as dead (see § A.3, *supra*). Such an upper brain definition of death would include not only adults in a persistent vegetative state but also anencephalic newborns, who are born with no upper brain. One could argue for treating the latter as dead even if upper brain death were not generally allowed, since the complete absence of the upper brain removes any doubt about whether their condition is reversible. But the only case on point held that such a baby was not dead under state law, and therefore could not be used as a source of retrievable organs, as its parents had hoped. In re T.A.C.P., 609 So.2d 588 (Fla. 1992).

Some suggest another halfway measure, short of full adoption of upper brain death: relax the "dead donor" requirement to allow organ retrieval from terminally ill or permanently unconscious patients who have specified, in an advance directive, that this

is their preference. Such patients may not see any detriment to their interests, and their consent provides assurance of their wishes. The problem is that since the patient is still alive under current law, removing an essential organ effectively constitutes murder—to which the victim's consent is no defense. The legal result is not altered by calling the procedure assisted suicide, as we see when that topic is explored in the next chapter. However, the potential to transplant more organs lends additional support to those who advocate legalizing voluntary euthanasia.

c. Mandates, and Novel Rules, for Organ Procurement

(1) After Death

Public surveys typically report broad support for organ donation, but the support does not necessarily translate into donation. Consent may not have been given by the individual before death, and family members may withhold surrogate consent after death. A number of commentators therefore have urged adoption of "presumed consent." Instead of assuming that people do not want to donate organs after death unless they consent (an "opt-in" policy), we might assume that people do want to donate organs unless they object (an "opt-out" policy). Many European countries have adopted presumed consent, and donation rates appear to be higher in those countries.

In addition, limited presumed consent statutes were common in the United States until 2007. For a few decades, states commonly authorized coroners or medical examiners to remove corneas from corpses during a state-authorized autopsy, so long as the coroner or medical examiner was unaware of an objection by the decedent or family. These laws typically imposed no concomitant duty to notify the family or ascertain their views, and few people knew they had the burden to object preemptively. Nevertheless, several courts upheld constitutional challenges to such rules, relying on the state's interest in increasing the supply of useable corneas and the relatively limited nature of the bodily intrusion. Two federal courts of appeal, however, held that the same state laws that give the family a right to object also confer on them a constitutionally cognizable "property" interest in the deceased's corneas, which carries with it some (unspecified) *procedural* due process protection (e.g., reviewing the decedent's medical record to see if an objection to donation had been lodged). It is this procedural

protection that is violated when a coroner fails to ascertain the family's wishes, even though there may be no *substantive* constitutional bar to the state more forthrightly allowing unconsented corneal harvesting. Brotherton v. Cleveland, 923 F.2d 477 (6th Cir. 1991); Newman v. Sathyavaglswaran, 287 F.3d 786 (9th Cir. 2002).

With encouragement from the 1987 version of the Uniform Anatomical Gift Act, many states also authorized coroners to remove vital organs, rather than just corneas, though they often required some

efforts to contact the family. If contacted, the family could preclude organ retrieval; if not, coroners could proceed, but they were generally reluctant to do so. Unlike the corneal transplant laws, these laws were not widely used, nor was their constitutionality tested on procedural or substantive grounds.

Policies for cornea or organ retrieval during state-authorized autopsies were largely reversed after promulgation of the 2006 version of the Uniform Anatomical Gift Act. Because of the concerns about process raised by cases like *Brotherton* and *Newman*, the drafters of the revised Act removed the provisions for unconsented retrieval of corneas, other tissues, or organs, and most states have followed the revised Act. Orentlicher, *Presumed Consent to Organ Donation: Its Rise and Fall in the United States*, 61 Rutgers L. Rev. 295 (2009).

An alternative approach, somewhat less radical, is to require all individuals to make an explicit living choice about donation. "Mandated choice" laws, under which all competent adults have to explicitly state whether or not they wish to be organ donors after

they die, are common. Almost all states ask applicants whether they want to be an organ donor when they obtain a driver's license. Supporters of mandated choice observe that it places the decision in the hands of the individual rather than in the hands of family members.

Critics worry that mandated choice usually is implemented without the prospective donor receiving much information, preventing genuine informed consent and causing many people to refuse consent.

While there has been an increase in the percentage of people who have registered as organ donors, mandated choice does not appear to have increased the number of organ transplants. Paula Chatterjee et al., *The Effect of State Policies on Organ Donation and Transplantation in the United States*, 175 JAMA Internal Med. 1323 (2015).

(2) During Life

Can organs not essential to a competent person's life ever be taken against the person's wishes? In the limited circumstances in which the question has arisen, the law has consistently prohibited such a step, even where the tissue is renewable and the procedure may be lifesaving. E.g., McFall v. Shimp, 10 Pa. D. & C3d 90 (1978) (discussed in Chapter 8.D.1). It is not permissible to risk one person's health even to save another person's life. Thus, for example, as discussed in Chapter 8, the government must permit abortion after fetal viability when necessary to protect the health of the pregnant woman. For an interesting argument defending the ethics of a law that would both compel everyone to be potential donors of renewable and non-essential tissues, and assure universal

access to such tissues by all donees in need, see G. Calabresi, *Do We Own Our Bodies?*, 1 Health Matrix 5 (1991).

Difficult problems arise where the question is whether to allow organs to be taken from incompetent, living persons for the benefit of others, a situation in which the capacities to form a donative intent and to understand the procedure, as well as

voluntariness, may all be lacking. There are two leading cases, with contrasting facts and results. Strunk v. Strunk, 445 S.W.2d 145 (Ky. 1969), upheld an order authorizing a kidney transplant from an intellectually disabled institutionalized man to his brother who was dying from kidney disease. No other family member had the requisite tissue compatibility, a cadaveric kidney apparently had not become available or was unsuitable, and the evidence indicated that the ward would be harmed less by the loss of a kidney than by the death of his brother, on whom he was quite emotionally dependent and who would be his caretaker after their parents died.

In contrast, In re Pescinski, 226 N.W.2d 180 (Wis. 1975), refused, ostensibly on jurisdictional grounds, to authorize transplant of a kidney from an institutionalized schizophrenic man to his sister for the same disease as in *Strunk*. Unlike *Strunk*, in *Pescinski* the incompetent brother's interests in being a donor were not as strong: he showed "marked indifference" to his environment, suggesting he had no real relationship with his siblings. Moreover, a competent brother was compatible but unwilling to donate. Thus, there was no way to avoid the conclusion that an incompetent person was being exploited for another's benefit. Chapter 8.B.4 discusses similar concerns—complicated by the countervailing issues of parental and reproductive freedom—raised by reports of

some parents who have conceived a child who, as an infant, can donate stem cells from blood or bone marrow to a sibling suffering from fatal anemia.

286

(3) Ownership of Organs

Underlying many of these debates over organ procurement strategies are assumptions or disagreements over who, if anyone, "owns" organs and other body tissues, either during or after life. By "own," we mean any number of particular property-like interests or rights, including the right to control, dispose of, and/or sell human tissues. Rather than declare any universal property status or immanent quality of human tissue, the law has reached a series of variable results based on particularized circumstances, informed by the value of a particular organ or tissue, the parties making claims upon it, and competing social interests. Space does not permit full discussion, but the range of possible results is suggested by the multiple opinions in the *Moore* case, discussed in Ch. 2.B.3.b, which posed ownership issues in a research, rather than a transplantation, context.

Moore held that a patient from whose tissues a commercial cell line was derived, without his knowledge or consent, had no tort claim for conversion, on the ground that he had no property interest in the cells once they were removed (with consent) from his body. (The cells came from a diseased spleen.) The majority was concerned that allowing a conversion action would unduly inhibit medical research. A concurring justice found it ethically abhorrent to recognize human tissue as a commodity in which one has a financial interest. A concurrence/dissent argued that a conversion claim *could* lie under proper facts, noting that the law

already recognizes property-like interests in controlling the disposition of one's organs. And a dissent argued strongly for immediate recognition of the conversion claim, noting that the defendants were asserting a property interest in the cell line. As these opinions, and the divergent results in other settings, suggest, there is much disagreement on the ownership issue. So far, however, courts have tended to follow the *Moore* court in rejecting any ownership rights for patients in tissues removed from their bodies for research, Greenberg v. Miami Children's Hospital Research Institute, 264 F. Supp. 2d 1064 (S.D. Fla. 2003); Wash. Univ. v. Catalona, 437 F. Supp. 2d 985 (E.D. Mo. 2006), although the *Greenberg* court did recognize the possibility of a claim of unjust enrichment by patients who were misled by the researcher. The law will continue to struggle with related debates, which are now at our doorstep, over ownership of new, bioengineered life forms.

2. ALLOCATION OF ORGANS

The National Organ Transplant Act of 1984 created the Organ Procurement and Transportation Network (OPTN), a private nonprofit entity, to oversee retrieval of organs and determine standards for their allocation. The United Network for Organ Sharing (UNOS) operates OPTN under contract with the Department of Health and Human Services (DHHS). UNOS is a membership organization that includes the country's 58 organ procurement organizations (OPOs), transplant centers, and others.

The first step to a needed transplant, from a patient's perspective, is getting on a waiting list. Once that occurs, the UNOS guidelines determine how OPOs allocate organs within their geographic area. We explore both steps.

a. Waiting Lists and UNOS Distribution

A prospective donee's eligibility for a waiting list is determined locally, by individual transplant centers (generally located at select major hospitals), which develop and apply their own criteria. Scarcity requires rationing, and because failure to qualify at this critical first step effectively terminates a candidate's eligibility, the stakes are very high. Many centers consider factors such as the prospects for successful surgery, the duration of the expected benefit, and the post-transplantation quality of life. Least controversial as bases for *excluding* candidates are medical considerations as to which there is some consensus, such as (in the case of a liver-transplant candidate) advanced heart or lung disease, current drug use or non-liver, non-skin cancer.

Other factors, though, have generated more ethical debate. For example, in the case of liver transplants, centers vary on when to list patients with a history of alcohol or other substance abuse; past abuse may have caused their disease (and the potential to resume their drug use may threaten the new organ). The decision may turn on sustained abstinence, social stability (e.g., employment, family support), and other factors. Some assert that while alcoholics should not be given a lower priority simply because

of their *alcoholism*, they should be held to account for failure to obtain *treatment* for their alcohol abuse that could have prevented their liver failure. Others think it is ill-advised to base standards on

notions of individual responsibility due to inadequate consensus about virtue, bias in deciding which behaviors will disqualify candidates for transplantation, and unjustified invasions of privacy. Some centers exclude felony prisoners, or struggle with eligibility of those who have recently attempted suicide. In one case, a 34-year-old woman with Down Syndrome was rejected for a heart and lung transplant by the medical centers at Stanford University and UC-San Diego because physicians thought she was not sufficiently intelligent to comply with the complicated post-transplant drug regimen. Though both institutions eventually relented, the case illuminates the problems inherent in such criteria. Decisions are also influenced by ability to pay and the adequacy of insurance coverage.

In recent years, concerns have been voiced about this system's overall fairness. The most serious arise from its "federated" nature: when an organ becomes available, preference is usually given to candidates in the local *geographic area* from which the organ came first, then to people within the larger *region* (of which there are 11 nationwide); and only last is it offered nationally. Allocating organs *within* these sets of nonoverlapping areas means that there can be significant differences in the waiting period—days to months, or more—*among* such areas, depending on supply and demand (and, sometimes, other factors). Regions with a sicker population, or whose medical

centers attract more transplant referrals, have longer wait lists, and patients are free to shop the country for areas with shorter lists, which are often those with less established transplant centers. This partly explains why baseball great Mickey Mantle received a liver

within a few days of acceptance on the waiting list, when the national average was two and half months.

The concern with fairness, along with the view that technology has diminished the justifications for localized use first (e.g., time limits on the useful life of a harvested organ), have led to calls for more uniform medical criteria for wait list decisions and a more nationalized organ distribution system. Some progress has been made in recent years, and various proposals are still being studied and debated. One proposal that would give priority to patients living within 500 miles of the donor's hospital may take effect in December 2020. Critics respond, in general, that these initiatives come at the urging of the large, established transplant centers in order to increase the allocation of organs they receive from other regions of the country with shorter wait lists and smaller centers. The larger centers reply, in part, that they should receive more organs because their patients are more seriously ill and they typically have a better track record than less experienced transplant teams.

b. The Ethics of Allocation Policies

Once an organ becomes available, a decision must be made about who it should go to. In addition to

geographic considerations, numerous allocation criteria are possible. "Social worth" is sometimes proposed, though it is widely rejected as unethical and unworkable in a pluralistic and humanistic society. Social worth may implicitly animate some decisions, however, perhaps in determining waiting list eligibility. Witness the following statement made by a lay member of one of the local committees assigned, in the 1960s, to allocate scarce kidney

dialysis machines: “I remember voting against a young woman who was a known prostitute. I found I couldn’t vote for her, rather than another candidate, a young wife and mother. I also voted against a young man who, until he learned he had renal failure, had been a ne’er-do-well, a real playboy.” The comparative valuation of human lives is of course ethically unacceptable as a general proposition, but we may deceive ourselves if we think that all such distinctions can be avoided. For instance, they necessarily come into play to some extent in decisions about medical prognosis.

Even when criteria are unobjectionable in the abstract, they are difficult to apply, or rank order, in practice. Fairness or equality among candidates suggests relying on the length of time one has been on the waiting list, but this could give the organ to someone who has the least need, for example because their physician is quick to add patients to the wait list. Maximizing the good that is done suggests giving organs to those who will live the longest, but this is in tension with another compelling principle known as the “rescue ethic,” which is to avoid the most imminent harms. Being fair, doing good, and

rescuing those in greatest peril are each valid principles, but they point to different patients, and people disagree on which should rank most highly. Therefore, present practice is to use a mixture of all three, with different weightings for different organs.

The OPOs, which are responsible for procuring organs in their region, make allocation decisions within regions based on uniform national policies promulgated by UNOS. While these policies differ among particular organs, typical criteria, to which specified weights are attached for ranking purposes, include:

- the medical urgency of the recipient's circumstances (this factor counts heavily in liver transplantation but is generally less critical for kidneys, where dialysis can often sustain the waiting donee);
- the likelihood of a successful transplant, including the expected length of the benefit (which in turn may depend, in part, on the recipient's age and general health status);
- biological compatibility between donor and recipient (e.g., organ size, blood type, and genetic makeup);
- how long the candidate has been on the waiting list; and
- the candidate's ability to be transplanted immediately.

Under the liver transplant criteria, for example, maximizing benefit and medical urgency both are

important, with the highest priority given to candidates with sudden liver failure who are expected to die within a week from their liver disease. (Patients with sudden liver failure and no pre-existing liver disease tend to do better with a transplant than do patients with chronic liver failure.) For most organs, there is a point system that specifies how to weight each factor in a particular case to generate a final ranking.

As the gap between supply and need for organs continues to increase, the fairness of allocation policies becomes even more important. Accordingly, UNOS regularly revisits its standards, and progress has been made in reducing the role of geographic considerations, narrowing racial differences in access to transplantation, and ensuring greater consistency in allocating organs from one center to another.

CHAPTER 7
END-OF-LIFE MEDICAL DECISIONS

As we saw in Chapter 2, patients possess the basic right of informed consent: Physicians and other health professionals may not provide care to a patient without first informing the patient about the care and its alternatives and obtaining the patient's voluntary and competent consent to the treatment. If the patient must consent to treatment, it follows that the patient also has a right to withhold consent and refuse the treatment.

Ordinarily, this corollary right to refuse treatment is not controversial. Patients with lower back pain from a slipped disk are free to choose among surgery to remove the disk or alternatives like anti-inflammatory drugs, exercise, and chiropractic manipulation. Patients with coronary artery disease may choose among invasive procedures, such as coronary artery bypass surgery and coronary angioplasty, or medical therapies, such as antianginal and cholesterol-lowering drugs.

In many cases, however, a refusal of treatment will result in the patient's death. In such cases, the state (or the health care provider) may want to invoke its interest in preserving life to ensure that the patient receives the treatment necessary to sustain life. The issue, then, is how we balance the individual's right to refuse treatment with the state's or other persons' interests in preserving the patient's life. To what extent does the right to refuse treatment include a right to refuse *life-sustaining* treatment?

As we will see, for patients who are able to make their own medical decisions, the law recognizes a broad right to refuse life-sustaining treatment. While early court decisions extended the right only to patients with serious illness or to especially burdensome treatments, the right now generally exists regardless of the patient's condition or the nature of the treatment. A young person in otherwise good health can refuse antibiotics for pneumonia just as an older person can refuse chemotherapy to treat an advanced and aggressive cancer.

A. REFUSAL OF LIFE-SUSTAINING TREATMENT

Legal recognition of the right to refuse life-sustaining treatment reflected the confluence of several factors. In recent decades, advances in medical technology have permitted physicians to save many lives that once were lost to disease or injury. In some cases, however, the person survives with a very poor quality of life and no hope of recovery. While medical care can maintain the person's life for weeks, months, or even years, the greatly diminished quality of life and the burdensomeness of the treatment mean to some individuals that the treatment is not desired.

In addition, these life-saving advances in technology have become very costly. If the life that can be saved has a very poor quality and society's limited resources could be used for other patients who would have a better quality of life and who want

treatment, then it is not clear that the state has an interest in always preserving patients' lives.

Finally, the law not only had recognized a right of informed consent, it also had recognized a right of pregnant women to

choose an abortion. In 1973, the U.S. Supreme Court handed down its decision in Roe v. Wade, 410 U.S. 113 (1973). If the state's interest in preserving the life of a fetus does not overcome the woman's interest in personal autonomy, then it becomes more difficult to justify a state interest in preserving the life of a dying person against the person's will.

In re Quinlan, 355 A.2d 647 (N.J.1976), was the first well-known modern case to consider whether physicians may intentionally withdraw or withhold life-sustaining medical treatment from a seriously ill patient. Only 22 and for reasons never established, Karen Quinlan fell into a coma, which evolved into a "persistent vegetative state." She retained some brainstem function, and therefore was not brain dead, but her physicians believed she would never regain consciousness and would soon die if they discontinued the use of a ventilator to maintain her breathing. When Quinlan's physicians declined to follow the instructions of her father, Joseph Quinlan, to discontinue the ventilation, he sought a court order appointing him guardian of his daughter, with express authority to end "all extraordinary medical procedures." Ultimately, the New Jersey Supreme Court held that the father, along with Quinlan's physicians (whom as guardian he could appoint or dismiss), could withdraw the ventilator after

confirming Quinlan's gloomy prognosis with a hospital committee.

But the court did not establish a broad right to refuse life-sustaining treatment. Rather, it held that the patient's right to make medical decisions should be balanced against the state's interests in preserving life and maintaining the integrity of the medical profession. According to the court, the interest in refusing

treatment grows "as the degree of bodily invasion increases and the prognosis dims." Because Quinlan's prognosis was "extremely poor" and the bodily invasion from treatment "very great," her rights prevailed over the state's interests.

After the *Quinlan* case, court decisions in New Jersey and other states refined the analysis of the right to refuse treatment. The courts considered the nature of the individual interest in refusing treatment and the range of state interests in providing treatment. The courts also settled on a broad view of the right to refuse treatment, one that generally is not qualified by the patient's prognosis or the burden of treatment.

1. THE INDIVIDUAL INTERESTS AT STAKE

Coloring almost all debates over the appropriate decisions in termination of care cases is the principle of personal autonomy: the belief that individuals have the right to make decisions as to matters that primarily affect themselves. Personal autonomy is an extraordinarily powerful principle in the modern western democracy. It is foundational to our political thinking; the vindication of personal autonomy is one

of the primary aspects of the historical movement from kings and dictators to constitutional democracies. It lies at the heart of many American constitutional principles.

"The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant." John Stuart Mill's views, expressed in *On Liberty* in 1859, strike most modern Americans as essentially

correct, even if they appreciate that governance is necessarily a bit more complicated than that. Our belief in the importance of personal autonomy is now almost instinctual.

Since we start from such strongly held commitments, it is hardly surprising that most people think it self-evident that people may not be compelled to submit to medical treatment against their will, at least when we have competent adults deciding for themselves. Nor is it surprising that most courts will seek to find, in the autonomy principle, warrant for patients' decisions on the termination of care.

In the early cases that recognized a right to refuse treatment, courts rested the right on two personal autonomy interests: the common-law right to be free of nonconsensual bodily invasion (i.e., the right to informed consent) and the substantive due process right to make decisions of critical importance to one's destiny (i.e., the right to privacy).

As the U.S. Supreme Court began to narrow the reach of the right to privacy in the 1980's, state courts relied more heavily on common-law principles of informed consent to find a right to refuse life-sustaining treatment. After the Court recognized a constitutional right to refuse treatment in the 1990 case, Cruzan v. Director, Missouri Department of Health, 497 U.S. 261 (1990), state courts resumed their reliance on substantive due process, although now framed per the *Cruzan* Court as an interest in liberty rather than privacy.

Although we can easily view the informed consent right and the substantive due process right as distinct rights—the first, a right to be free of unwanted bodily invasion; the second, a right to make important personal decisions more generally—the *Cruzan* Court

collapsed the two. There, the Court turned to common-law principles of informed consent as the basis for finding a constitutional right to refuse life-sustaining treatment.

2. THE STATE'S INTERESTS

While the *Quinlan* Court discussed the two state interests in preserving life and the integrity of the medical profession, later courts generally identified four state interests in requiring the administration of life-sustaining treatment.

As these courts observed, the state has important interests in preserving life, maintaining the integrity of the medical profession, preventing suicide, and protecting the welfare of innocent third parties. Each

interest might justify limits on the patient's freedom to refuse unwanted treatment.

The Nevada Supreme Court identified a fifth state interest —"encouraging the charitable and humane care of afflicted persons." McKay v. Bergstedt, 801 P.2d 617, 628 (Nev. 1990). Patients contemplating refusal of life-sustaining treatment therefore must be fully informed of the care alternatives that would be available to them if they remained alive. The case of Larry McAfee illustrates the importance of this interest. After a motorcycle accident that left him quadriplegic and ventilator dependent. McAfee sought removal of his ventilator, and the Georgia Supreme Court recognized his right to refuse treatment. State v. McAfee, 385 S.E.2d 651 (Ga. 1989). However, he did not exercise his court-authorized right, in part because wide publicity about his case

brought forth support services that made his life more worthwhile to him

3. BALANCING THE DIFFERENT INTERESTS

As discussed, the *Quinlan* Court limited the right to refuse treatment based on the severity of the patient's condition and the burdens imposed by treatment. In subsequent cases, the New Jersey Supreme Court and other courts generally extended the right to refuse treatment to all patients and all treatments.

According to the later court decisions, the state's indirect and abstract interest in preserving the life of the patient must yield to the much stronger personal

interest of patients in directing the course of their own lives. As the California Supreme Court wrote,

> For self-determination to have any meaning, it cannot be subject to the scrutiny of anyone else's conscience or sensibilities. It is the individual who must live or die with the course of treatment chosen or rejected, not the state. Thor v. Superior Ct., 855 P.2d 375, 384 (Cal. 1993).

Still, the *Cruzan* Court left open the possibility that the right to refuse treatment could be limited based on the patient's medical condition, writing that "We do not think a State is required to remain neutral in the face of an informed and voluntary decision by a physically able adult to starve to death." 497 U.S. at 280. Similarly, in the *Bergstedt* case, the Nevada Supreme Court limited the right to refuse treatment to patients who are "irreversibly sustained or subject to being sustained by artificial life support systems." 801 P.2d at 624.

While *Cruzan* and *Bergstedt* suggest that the state's interest in preserving the patient's life might restrict the right to refuse treatment, courts have not imposed any practical limits on the right to refuse for competent adults. When competent patients refuse life-sustaining treatment, they typically do so because of irreversible illness. It is a rare case in which a competent patient refuses treatment that will restore good health, other than when treatment would violate religious beliefs. Hence, only in some very early cases did courts override refusals of treatment in order to promote the state interest in preserving life, and neither the U.S. Supreme Court

nor any state supreme court has mandated the provision of life-sustaining treatment over the objections of a competent adult, absent the kind of third party interests discussed below.

Allowing refusals of treatment also would not compromise the integrity of the medical profession since the profession recognizes that at some point, treatment should turn from trying to cure the curable to trying to ensure a comfortable dying process for the incurable. *Quinlan*, at 355 A.2d at 667.

As to the state's interest in preventing suicide, we will explore that interest at greater length in the section on aid in dying. For now, the key point is that courts have distinguished between patients who die from the natural progression of their terminal illness and patients who bring on their death actively, as with a lethal dose of medication.

Finally, third party interests can limit a right to refuse treatment, but not in the typical end-of-life case. Third-party interests arise primarily in the contexts of public health or pregnancy. As to the

former, courts permit a number of treatments to be imposed when a patient's illness threatens the health of others, as with mandatory immunizations or antibiotics for communicable diseases. The classic case here is Jacobson v. Massachusetts, 197 U.S. 11 (1905), in which the Supreme Court rejected a challenge to mandatory immunization for smallpox. With regard to pregnancy, some courts will prohibit a woman from refusing treatment, such as a blood transfusion or cesarean section, that is necessary to sustain the life (or health) of both her fetus and

herself. These cases will be discussed in Chapter 8. There also are cases involving prisoners in which the court has ordered treatment because of concerns about the disruption of institutional order.[1] But in the usual end-of-life case, when relevant third party interests are not at stake, the patient may refuse any and all treatments.

The generally unqualified nature of the right reflects a number of considerations. As we saw, the *Quinlan* court considered the severity of the patient's condition in deciding whether the state's interests could override the patient's right to refuse. But it is difficult to draw lines in this area. When, for example, does a patient's condition cross the line from hopeful to hopeless? When the chances of recovery are 10 percent, 5 percent, less than one percent? As a California court of appeals wrote in the *Bouvia* case,

> As in all matters, lines must be drawn at some point, somewhere, but that decision must ultimately belong to the one whose life is in issue. . . . It is not a medical decision for her physicians to make. Neither is it a legal question whose soundness is to be resolved by lawyers or judges. . . . It is a moral and philosophical decision that, being a competent adult,

is hers alone. Bouvia v. Superior Court, 225 Cal. Rptr. 297, 304–305 (Ct. App. 1986).

305

Also, on matters involving the quality of life, it is dangerous to give government the power to decide whose lives must be preserved and whose need not be. The state is likely to exercise such a power in ways that favor some patients and disfavor others based on race, sex, income, or other inappropriate factors.

Allowing patients to decide for themselves need not put patient welfare at undue risk. As mentioned, mentally competent people typically do not refuse life-sustaining treatment unless they are suffering from irreversible illness. And when they do, it is usually on account of religious scruples, as with Jehovah's Witnesses and blood transfusions.

After recognition of the right to refuse life-sustaining treatment proceeded gradually from state to state, the U.S. Supreme Court weighed in. Fourteen years after the *Quinlan* decision, in *Cruzan*, the Supreme Court concluded that the right to refuse life-sustaining treatment was protected by the Due Process Clause. Nancy Cruzan's physicians had placed a feeding tube down her throat when she lapsed into an unconscious state three weeks after suffering injuries in an automobile accident. Later, they concluded that their patient, like Karen Quinlan, was in a persistent vegetative state. When hospital personnel refused to honor Cruzan's parents' request to withdraw the artificially provided nutrition and hydration, her parents sought a court order. On appeal, the U.S. Supreme Court recognized the right to refuse life-sustaining treatment as a constitutional right.

Technically, the *Cruzan* majority only assumed for purposes of the case that individuals enjoy a constitutional right to refuse life-sustaining treatment. Nevertheless, the decision has been read by courts and commentators as establishing such a right, and the Supreme Court effectively approved of that reading in its *Glucksberg* decision on aid in dying.

In the second part of the case, which we will take up later, the Court considered how the right to refuse treatment can be exercised for patients who lack the intellectual capacity to make medical decisions for themselves. For such patients, a number of questions arise:

1. If we know patients' views about medical care, must we necessarily follow them? Or should we decide based on our sense of the patient's best interests?

2. If the patient's views govern, what evidence do we need to establish what those views are? Karen Quinlan and Nancy Cruzan could not themselves instruct their physicians. There was testimony that Quinlan had expressed, in social conversations, a distaste for heroic medical measures. Are such reports adequate to establish what she would now want done? Do even fully documented advance directives deserve the same deference as the patient's contemporaneous instructions?

3. Who speaks on behalf of a patient who leaves no guidance? Is a family relationship sufficient to confer authority, or is a court appointment needed?

What standard is appropriate to guide the decisionmaker's judgment?

We will consider each of these issues, and others as well. But before doing so, we will examine some of the ethical considerations that once were invoked to distinguish permissible withdrawals of care from impermissible ones.

4. POTENTIAL ETHICAL DISTINCTIONS

a. Withholding Versus Withdrawing Treatment

At one time, some commentators argued that it was permissible to *withhold* life-sustaining treatment but not to *withdraw* such treatment. In this view, a patient could refuse the initiation of artificial ventilation, but not its discontinuation once it was started. And this would have been analogous to rules about the duty to come to the aid of a person in distress. While there is no obligation to come to someone's rescue (a withholding of aid), there is an obligation not to abandon a rescue (a withdrawal of aid).

Over time, however, in both ethics and the law, the distinction between withdrawing and withholding treatment was rejected. Partly this reflected the overlap between the two categories. One can withdraw artificial feeding by removing a patient's feeding tube, and one can withhold artificial feeding by not connecting the next container of sustenance to the feeding tube. Moreover, as many scholars have observed, it is arguably worse to withhold than to withdraw. If treatment is withheld, then an

opportunity is lost to see if the treatment would provide unexpected benefit. Withdrawal presumably occurs only after it becomes clear that the treatment provides insufficient benefit. In addition, as

suggested by Justice William Brennan in his dissent in the *Cruzan* case, if we did not recognize a right to have treatment withdrawn, some people might not seek care in the first place because they would be afraid of not being able to stop treatment once it was started.

Despite years of ethicists and courts rejecting the distinction between withholding and withdrawing, health care providers may feel very differently about the two acts. Some physicians believe it is less acceptable to withdraw care than to withhold it. These feelings are not surprising. It is hard not to feel responsible for a patient's death when you turn off a ventilator and the patient dies within minutes. Accordingly, while patients have a right to refuse treatment, health care providers generally can decline to participate in the withdrawal of care and arrange for other providers who are comfortable with the withdrawal to carry out the patient's wishes.

b. Type of Treatment at Stake

As mentioned, while the *Quinlan* Court considered the burdens of treatment when defining the extent of the right to refuse treatment, later courts abandoned that analysis. It no longer matters whether a patient needs a ventilator and other highly invasive intensive care or a simple blood transfusion.

But until the *Cruzan* decision, there was considerable debate on the question whether the right to refuse treatment should include artificial nutrition and hydration. As Daniel Callahan wrote,

> The feeding of the hungry, whether because they are poor or because they are physically unable to feed themselves, is the most fundamental of all human relationships. . . . It is a most

dangerous business to tamper with, or adulterate, so enduring and central a moral emotion. 13(5) Hastings Center Rep. 22 (1983).

On the other hand, an important argument for treating nasogastric tubes and the like as one treats other treatment draws on the analogy to artificial ventilation. Food and water are hardly more basic to human life than air. Yet while it is obvious that no court or ethicist would ever consider confining a patient in an airtight room to die of suffocation, it has become routine to withdraw ventilators from patients unable to breathe on their own. Food, water, and air are not themselves medical treatments, but the method of delivering them is when patients are unable to breathe or ingest on their own. The *Cruzan* case essentially resolved the debate in terms of the law. As Justice Sandra Day O'Connor wrote in her concurring opinion, "Artificial feeding cannot readily be distinguished from other forms of medical treatment" because artificial feeding inevitably involves "some degree of intrusion and restraint." 497 U.S. at 288. It is now widely accepted that patients can refuse artificial nutrition and hydration as they can any other medical treatment.

Nevertheless, there are times when courts treat artificial nutrition and hydration differently. As we will discuss below, an important question is how one decides whether to exercise an incompetent patient's right to have life-sustaining treatment withdrawn. And in that context, the nature of the treatment can matter. For example, after the Missouri Supreme Court held that Nancy Cruzan's feeding tube could not be withdrawn absent "clear and convincing" evidence of her wishes not to be artificially fed, a Missouri court of appeals held that the *Cruzan* standard did not necessarily apply to

other kinds of treatment. In In re Warren, 858 S.W.2d 263 (Mo. Ct. Appl. 1993), a guardian agreed to a do-not-resuscitate (DNR) order for a permanently unconscious woman. The court held that even though there was no evidence of Julia Warren's wishes, the guardian could decline resuscitation (CPR) on the basis of his general duty to make health care decisions in terms of Warren's best interests.

As the *Warren* case indicates, an important question is how much significance to attach to the fact that many of the key end-of-life cases involved withdrawal of artificial nutrition and hydration, rather than artificial ventilation, dialysis, or surgery. As one reads the opinions, it generally does not appear that the kind of treatment matters. But it is difficult to believe that judges would be as strict if the treatment at stake were surgery to replace a heart valve. The distinction between artificial nutrition and hydration and other treatments extends beyond judges. Many health care providers are still slower to withdraw feeding tubes than stop other treatments

because of their personal moral concerns. Family members also may be more reluctant to authorize withdrawal of nutrition and hydration. When Karen Quinlan's father requested withdrawal of her ventilator, he still wanted to maintain her feeding. And while many of the early end-of-life cases involved artificial ventilation, the more recent cases involve artificial nutrition and hydration, as in the case of Terri Schiavo.

In that case, a dispute between Schiavo's husband and parents resulted in nearly seven years of litigation with more than a dozen state and federal court decisions, statutes by the Florida Legislature and Congress, and intervention by Governor Jeb Bush and

President George W. Bush. Schiavo's husband requested the withdrawal of her feeding tube after she had been in a persistent vegetative state for several years. After reviewing the testimony, the trial court judge concluded that there was clear and convincing evidence that Schiavo would not want a feeding tube, based on her prior oral statements to family members. In re Schiavo, 780 So. 2d 176, 179–180 (Fla. Ct. App. 2001). Her parents petitioned successfully for reinsertion of the feeding tube, and after a second removal and reinsertion in 2003, Schiavo's feeding tube was finally removed in March 2005. Her death ensued 13 days later.

The prolonged proceedings in the *Schiavo* case not only reflected the fact that feeding was at stake. It also was important that the dispute was among family members rather than between family and

health care providers. Family disagreements are an important reason for cases to end up in court.

To an important extent, this issue is addressed by state statutes governing surrogate decisionmaking. Typically, these statutes establish a hierarchy of family members who have the authority to make decisions on behalf of the patient, with spouses having first priority, then adult children, parents, and adult siblings. A number of states also include other relatives or close friends as potential surrogate decisionmakers. However, not all states have such statutes, and even when they exist, physicians may be reluctant to implement a request by some family members to discontinue treatment over the objections of other family members. Courts also are reluctant to permit withdrawal of treatment in the presence of a disagreement among family members.

While the right to refuse artificial nutrition and hydration is firmly established, debate persists as to whether a person can refuse food and water that can be ingested naturally. That debate will be considered in the section on aid in dying.

* * *

As we have seen, decisions whether to provide medical treatment generally must be left to the patient. The fundamental importance of personal autonomy has led to a broad right to refuse life-sustaining treatment.

The principle of personal autonomy is of course most obviously relevant to the case of alert,

competent adults capable of expressing their views at the time of the treatment decision.

To what extent does the principle of personal autonomy apply to patients who no longer possess decision making capacity? We might be guided by their previously expressed wishes. But what if those wishes are ambiguous? Or what if they never expressed their wishes? And what about children or other patients who never possessed decision making capacity?

We will consider each of these scenarios. First, though, we will consider how decision making capacity is determined.

5. THE PATIENT WHOSE COMPETENCE IS UNCERTAIN

In the vast majority of cases, it is obvious whether a patient is competent. Most patients either clearly possess decision making capacity or they clearly do not. In some cases, however, whether the patient can competently consent to or refuse treatment is not so easily decided. While patient competence has been a long-standing

issue, physicians, lawyers, and other professionals have not yet developed a readily applied standard for assessing competence, perhaps because of its elusive nature. It is not a simple matter to assess a person's capacity for reasoning and understanding, nor is it clear what level of understanding a person must be able to exercise to be considered competent.

314

The issue is further complicated by the fact that people can possess decision-making capacity for some matters but not others. Whether a patient possesses adequate decision-making capacity must be assessed in relation to the specific decision to be made, and not by global assessments of the person. Thus, for example, a legal judgment (in either direction) of whether a person is competent to manage their financial affairs or the logistics of daily living is not necessarily dispositive on the separate question of whether a person has the capacity to make a particular health care decision.

How then is capacity to make health care decisions to be judged? Important considerations include whether the patient's decision is based on "rational" reasons and whether the patient has the ability to understand or has demonstrated actual understanding. For example, the New Jersey Supreme Court has written that "[a] competent patient has a clear understanding of the nature of his or her illness and prognosis, and of the risks and benefits of the proposed treatment, and has the capacity to reason and make judgments about that information." In re Farrell, 529 A.2d 404, 413 n.7 (N.J. 1987).

Note that while the decision needs to be based on reasons, it does not have to be reasonable in the eyes of the family, physician, or a court. That a patient's decision may seem unwise is not sufficient to

override it. The patient must be acting irrationally or without understanding.

315

In other words, the patient's capacity is not appropriately judged by reference solely to the content of the patient's decision. We cannot conclude that a patient lacks the necessary decision-making capacity just because the attending physicians believe that the patient's decision is "wrong" or "unwise."

At the same time, it is clear that the content of a patient's decision will influence the assessment of capacity to some degree. At a practical level, it is inevitable that patient decisions which doctors and family view as reasonable will rarely if ever be challenged for incapacity. Even if a medical expert or court were to conclude that the patient lacked capacity, it would not change the outcome. Under a best-interests-of-the-patient standard, the decisions would come out as the patient indicated.

And when a decision seems unreasonable, the content of the decision also cannot be entirely irrelevant to the assessment of capacity. Especially because judgments of capacity will be uncertain in some cases, it becomes important to determine how convinced we must be of patients' capacity before respecting their choices. We need to limit the risk that a patient refusing treatment would incorrectly be deemed to possess decision-making capacity

And this is a significant risk. In a review-of-the-literature study, researchers found that physicians "missed the diagnosis in 58 percent of patients who were judged incapable" in a formal, independent assessment. In other words, in more than half the patients who lacked decision-making capacity, the

physicians incorrectly concluded that the patients possessed such capacity. On the other hand, physicians generally were correct when they made a diagnosis of incapacity. Laura L. Sessums, *Does This Patient Have Medical Decision-Making Capacity?*, 306 JAMA 420 (2011).

Given the risks of a mistaken finding of capacity, many experts believe that we may demand clearer evidence of the patient's decision-making capacity as the consequences of following the patient's expressed preference become more substantial. For decisions that will end a patient's life, we would want to be especially confident that the patient has decision-making capacity.

Consider in this regard the case of Rosaria Candura, a 77-year-old woman with gangrene in her right foot and lower leg. Her physicians recommended amputation, but she refused after originally agreeing to the surgery. She already had undergone two amputations on her right foot, losing a toe in one and part of her foot in the other. In explaining her reasons for refusing a third amputation, she said that

> she has been unhappy since the death of her husband [two years earlier]; that she does not wish to be a burden to her children; that she does not believe that the operation will cure her; that she does not wish to live as an invalid or in a nursing home; and that she does not fear death but welcomes it. Lane v. Candura, 376 N.E.2d 1232 (Mass. App. Ct. 1978).

The court noted that Candura was lucid on some matters and confused on others. Nevertheless, the court held that she was

competent to refuse treatment, observing that her competence had not been questioned until she withdrew her original consent to the surgery and that she had a right to make her own decisions about medical treatment, even if the decisions seemed unwise. In addition, while she might have exhibited symptoms of senility, there was no evidence that her confusion or forgetfulness was interfering with her ability to decide about the surgery. Finally, her case did not involve "the uninformed decision of a person incapable of appreciating the nature and consequences of her act."

But query whether her unhappiness since her husband's death indicated that she was mentally depressed and that her refusal of treatment signaled the need for psychiatric intervention.

We also should consider whether a patient's refusal of treatment might reflect a mistaken assessment of what the future will bring. The *Bergstedt* case mentioned above provides a useful example.

Kenneth Bergstedt had been quadriplegic since a swimming pool accident at the age of 10. Although permanently dependent upon a ventilator, his condition was not terminal. He completed high school, and wrote poetry on a computer, read, and watched television.

His desire to continue with treatment changed at age 31. He had been dependent upon his parents since his accident, but his mother had died, and his father was fatally ill with lung cancer. Bergstedt, wrote the court, "despaired over the prospect of life without the attentive care, companionship and love of his devoted father." 801 P.2d at 620. One month before the Nevada Supreme Court upheld Bergstedt's right to refuse treatment, he died, apparently because his father disconnected the ventilator.

If Bergstedt had been alive when the court decided the case, it might have made sense for the court to require that he see what his life was like after his father died before deciding whether to refuse the ventilator.

While this makes sense, courts have not taken this approach, perhaps because of the difficulties with it. For example, in some sense, all refusals of treatment involve patient fears about the future —patients worry that their quality of life will be unbearable. And we could say that patients might change their mind over time as they find that the quality of their life is not as bad as expected. Indeed, recall that Larry McAfee, whose medical condition was similar to that of Bergstedt's, changed his mind after the Georgia Supreme Court recognized his right to refuse treatment.

But one of the important aspects of personal autonomy is that people should be free to decide for themselves how to weigh future uncertainties. Also, if the law required patients to give treatment a try for some period of time before approving refusals,

how long should the period last? Should Bergstedt have had to try life without a parent for one month, six months, a year?

a. Adolescents

Issues of competence also arise for adolescents. There are two separate questions that must be answered: (1) Do minors have capacity to decide for themselves? (2) If not, is the decision one that parents can make for them? In general, the answer to (1) is "no" and to (2) is "yes": Minors usually are held to lack decision-making capacity, and parents usually have authority to make decisions on their behalf.

Both rules have important exceptions, however. The exceptions to the second rule are discussed later in this chapter, where we learn that courts often deny parents authority to refuse life-sustaining treatment for their children. This section discusses exceptions to the first rule, that is, situations where minors may make important medical decisions for themselves.

While minors generally lack decision-making capacity, "mature minors" may be accorded decision-making authority if they can show decision-making capacity despite their age. The Supreme Court of Appeals of West Virginia analyzed the issue when the question of withholding ventilation arose for Larry Belcher, a minor of 17 years and 8 months. He suffered from muscular dystrophy, and it appeared that he would become ventilator dependent imminently. His physician and parents decided that he would not be ventilated or resuscitated in the event of a respiratory arrest, and he died the

following day from a respiratory arrest. A question before the court was whether his physician should have consulted him before a decision was made about ventilation and resuscitation. The court wrote:

> Whether the child has the capacity to consent depends upon the age, ability, experience, education, training, and degree of maturity or judgment obtained by the child . . . The factual determination would also involve whether the minor has the capacity to appreciate the nature, risks, and consequences of the medical procedure to be performed, or the treatment to be administered or withheld. Belcher v. Charleston Area Medical Center, 422 S.E.2d 827, 838 (W. Va. 1992).

The issue also came before the Maine Supreme Court in a case involving the withholding of a feeding tube from Chad Swan. He was 18 years old at the time of the decision but had become permanently unconscious from an automobile accident at age 17. In permitting the withholding, the court relied on statements that Swan had made at ages 16 and 17. According to the court, "capacity exists when the minor has the ability of the average person to understand and weigh the risks and benefits." In re Swan, 569 A.2d 1202, 1205 (Me. 1990).

Not all states recognize a mature minor doctrine. For example, a federal district court in Georgia concluded that Georgia state law grants decision-making capacity only in certain statutorily specified situations (e.g., minors who are married, pregnant, or have children) but does not include a mature

minor doctrine. Novak v. Cobb County-Kennestone Hospital Authority, 849 F. Supp. 1559, 1575–1576 (N.D. Ga. 1994). Other states have not decided one way or another whether maturity serves as the basis for decision-making capacity for a minor. In re Cassandra C., 112 A.3d 158 (Conn. 2015); In re Conner, 140 P.3d 1167 (Or. Ct. App. 2006). Many states authorize minors to agree to treatment in specific, sensitive circumstances, such as for substance abuse, mental health services, or sexually transmitted diseases. And the U.S. Supreme Court has recognized a right for mature minors to choose abortion without notifying their parents.

For more discussion, see Hall, Orentlicher, Bobinski, et al., Health Care Law and Ethics (9th ed. 2018) and Meisel and Cerminara, The Right to Die (3rd ed. 2010).

6. PATIENTS LACKING DECISIONMAKING CAPACITY

When a patient lacks decision-making capacity, as is common when decisions about life-sustaining treatment are made, a few questions arise: Does the right to refuse life-sustaining treatment survive incompetence? If it does, is the right the same? How is the right invoked on behalf of the patient?

As we will see, while it is clear that the right to refuse treatment survives incapacity, courts have varied with regard to the procedural requirements that must be met before life-sustaining treatment is withheld or withdrawn from incompetent patients.

a. Incapacity and the Right to Have Treatment Withheld or Withdrawn

Courts have consistently held that an adult's right to refuse life-sustaining medical treatment remains intact when the person loses decision-making capacity. As the Massachusetts Supreme Judicial Court wrote, the right "must extend to the case of an incompetent, as well as a competent, patient because the value of human dignity extends to both." Superintendent of Belchertown State School v. Saikewicz, 370 N.E.2d 417, 427 (Mass. 1977).

But to the extent that the right to refuse life-sustaining treatment rests on principles of self-determination, is it so clear that the right should exist for individuals who no longer can express their preferences?

Even though incapacitated patients cannot exercise self-determination, there still are important interests at stake. For example, we don't want individuals to refrain from starting treatment when they possess decision-making capacity for fear that their wishes to discontinue treatment later will not be respected. In addition, rights of self-determination reflect not only the

importance of individual control over key decisions but also the importance of not having government control over these decisions. Even if patients cannot make the decisions for themselves, we still do not want the government to make the decisions for them. Finally, incapacitated persons may not have interests in self-determination, but they do have interests in being spared from suffering and in being treated in a respectful manner.

323

A right to refuse treatment is needed to protect patients from unjustified suffering and disrespectful care.

b. Variation in Procedural Rules

The recognition of a right to refuse treatment for incompetent persons still leaves the question as to how the right should be implemented. In the *Cruzan* case, the U.S. Supreme Court gave states broad leeway to adopt procedural rules for deciding when life-sustaining treatment can be withdrawn from incompetent persons.

When Cruzan's parents sought to have her disconnected from the tubes that provided her nutrition and hydration, the Missouri Supreme Court required the doctors to honor this request only if there was clear and convincing evidence that it was *Cruzan*'s wish —a standard the court held was unmet. Affirming, the U.S. Supreme Court found nothing in the Constitution that required Missouri to apply a more lax evidentiary standard.

Even more importantly, the Court made clear that the Due Process Clause does not confer the right to refuse treatment on "anyone but the patient herself." Legislators might conclude it wise to give decision-making authority to close relatives in such cases,

but family members have no *constitutional* claim to act for the patient, for "there is no automatic assurance that the view of close family members will necessarily be the same as the patient's would have been had she been confronted with the prospect of her situation while competent." 497 U.S. at 286.

Given the importance of the issues at stake—literally life and death—the Court held that states could err on the side of preserving life and require clear and convincing evidence that the patient would want treatment withheld or withdrawn before discontinuing treatment.

A case from New York illustrates the concerns. A New York trial judge issued an order allowing the removal of a feeding tube from an 86-year-old stroke victim who had been unconscious for four-and-a-half months, and whom doctors had concluded was in a permanent vegetative state, but whose own views had not been clearly established. Over the weekend following the issuance of the order, but before the doctors acted, the woman awoke. Her physician was called to her bedside, found her alert, and proceeded to describe her legal case to her. She indicated she understood, and the physician then asked her what she wanted done. "She replied 'These are difficult decisions' and lapsed back into sleep." The judge then withdrew his order. N.Y. Times, April 15, 1989, at pg. A15.

While the *Cruzan* Court upheld Missouri's strict requirement of clear and convincing evidence, it did not require similarly strict procedural standards in other states. Across the different states, there is both overlap and divergence in their approaches.

There is virtue to allowing variation among the states. Everyone agrees that not every available medical procedure should forever be provided to an incompetent patient who left no directions and appointed no agent to decline them. The difficulty lies

in establishing the process for making decisions on behalf of such patients and the precise boundaries within which the treatment may be withdrawn. There are many factual nuances and difficult policy choices that must be made in working out such details, in part because the rules one adopts may affect physician and hospital behavior in ways that will impact other patients not before the court. One's policy judgments may also be affected by the often rapid changes in medical knowledge and medical technology. Much can be learned from the experiences with the varying solutions chosen by different state courts and legislatures. Invoking the Constitution to impose a uniform national rule that can be changed only by decisions of the U.S. Supreme Court, serving in effect as a national board of bioethics in this field, would short circuit the efforts by states to test different approaches.

While different approaches have emerged, there also is some similarity. All states permit withdrawal of treatment for patients who have left clear and convincing evidence of their treatment preferences—and doing so is probably required by the constitutional right to refuse treatment. A helpful definition of the standard can be found in a New York case, for which the court observed that evidence of the patient's wishes is clear and convincing when it is "sufficient to persuade the trier of fact that the patient had a firm and settled commitment" to decline treatment. In re Westchester County (O'Connor), 531 N.E.2d 607, 613 (N.Y. 1988).

In determining whether the patient provided sufficiently clear evidence, a range of evidence is relevant. Most probative would be a "living will" or other advance directives in which patients document their preferences. Often patients will have discussed their wishes with family members, friends, or health care providers in conversation, and courts will consider the patients' oral statements. Other considerations include religious beliefs, reactions that the patient voiced regarding medical treatment for others, or the patient's past practices with respect to treatment decisions.

Of course, the different kinds of evidence will vary in their reliability. Courts have observed that the probative value of evidence will depend on the remoteness, consistency, and thoughtfulness of prior statements or actions and the maturity of the person at the time of the statements or acts.

> Thus, for example, an offhand remark about not wanting to live under certain circumstances made by a person when young and in the peak of health would not in itself constitute clear proof 20 years later that he would want life-sustaining treatment withheld under those circumstances. In contrast, a carefully considered position, especially if written, that a person had maintained over a number of years or that he had acted upon in comparable circumstances might be clear evidence of his intent. In re Conroy, 486 A.2d 1209, 1230 (N.J. 1985).

The specificity of prior statements also can be important. Indeed, some courts have not found clear

and convincing evidence when oral statements did not speak specifically to the patient's condition. So, for example, having said, "I would never want to live in a nursing home," is very different from having said, "I would not want artificial feeding if I become permanently unconscious like Nancy Cruzan."

With the different factors, it is not surprising that courts have varied on whether past statements by patients constitute clear and convincing evidence.

For example, in the *Swan* case, the Maine Supreme Court thought there was clear and convincing evidence from two prior statements by Swan, who was permanently unconscious. The first was made during a discussion with his mother about a prominent case in Maine involving another permanently unconscious patient. Swan and his mother "discussed what it meant to be a 'vegetable,' " and Swan said that " 'If I can't be myself . . . no way . . . let me go to sleep.' " Swan's second statement occurred only eight days before the accident that left him permanently unconscious after he visited a family friend who was comatose after a car accident. According to his brother, Swan said " 'I don't ever want to get like that. . . . I would want somebody to let me leave—to go in peace.' " *Swan*, 569 A.2d at 1205.

On the other hand, in the *Jobes* case, the New Jersey Supreme Court did not find clear and convincing evidence from similar statements made by Nancy Jobes before an accident that left her permanently unconscious. She had told a life-long friend that "if she were ever crippled like the children

with multiple sclerosis and muscular dystrophy" that the friend cared for, "she would not want to live." Just a few months before

Jobes' accident, the friend also recalled Jobes saying that "she would not want to be kept alive on a respirator like a patient suffering from amyotrophic lateral sclerosis" whom the friend described. During the same time period, another long-time friend testified that Jobes "had told her that 'it was a shame that [Karen Quinlan] hadn't died when they removed the respirator; that that wasn't living, it was existing; that she had wished that God had taken her then. . . .' " Jobes' husband "generally recalled her having stated that she would not want to be kept alive under Karen Quinlan's circumstances. She did this frequently when the Quinlan case was in the news" (which was a few years before Jobes' accident). 529 A.2d 434, 442 (N.J. 1987).

If clear and convincing evidence does not exist, there is considerable variation from state to state as to whether or when treatment may be discontinued.

Substituted judgment. Courts and legislatures often permit family members or guardians to decide on the patient's behalf under a "substituted judgment" standard in which the decision-makers draw on their knowledge of the patient to reach "as much as possible the decision that the incompetent patient would make if he or she were competent." *Jobes*, 529 A.2d at 444. As the *Jobes* court observed,

> Almost invariably the patient's family has an intimate understanding of the patient's medical attitudes and general world view and therefore

> is in the best position to know the motives and considerations that would control the patient's medical decisions. Id. at 445.

And even when the family might not have a firm sense of what the patient would want, these courts and legislatures favor

substituted judgment based on a combination of how the family thinks the patient would decide and the patient's best interests. As the Indiana Supreme Court wrote:

> In our society, health care decision-making for patients typically transfers upon incompetence to the patient's family. . . . Even when they have not left formal advance directives or expressed particular opinions about life-sustaining medical treatment, most Americans want the decisions about their care, upon their incapacity, to be made for them by family and physician, rather than by strangers or by government. In re Lawrence, 579 N.E.2d 32, 39 (Ind. 1991).

State statutes conferring decision-making authority on family members are common, and they provide a simple approach when patients do not leave clear evidence of their wishes.

But there are limitations to these statutes. For example, they may apply only in the context of a terminal illness or permanent unconsciousness. In addition, the statutes vary in their definition of "family." Some include only spouses, parents, adult children, and adult siblings, while others include a broader range of blood relatives. Some states also include close friends or even any "adult who has

exhibited special concern for the patient, who is familiar with the patient's personal values." 18–A Me. Rev. Stat. Ann. § 5–805(c).

Best interests. Some courts have chosen a simple best interests standard for cases in which the patient's preferences cannot be established. Rasmussen v. Fleming, 741 P.2d 674, 689 (Ariz. 1987); In re Torres, 357 N.W.2d 332, 338–339 (Minn. 1984).

Under a best interests analysis, decisionmakers balance the benefits and detriments of treatment to see if the advantages outweigh the disadvantages. Relevant considerations include the extent to which treatment can extend life, the likelihood that treatment will be effective, and the quality of life both with and without treatment.

The best interests standard is particularly important for patient who never possessed decision-making capacity. It would not be possible to employ a true substituted judgment approach with patients who have never possessed decision-making capacity. How can one ask what the patient would have wanted, when by definition the patient has never had the capacity to form a view entitled to such deference? It would therefore seem that the only basis for withdrawing life-sustaining treatment in such cases would be an objective judgment that continuation of the treatment would not be in the patient's interests.

A good illustration of the best interests standard for the never competent patient is the Massachusetts

Supreme Judicial Court's discussion in the *Saikewicz* case. According to the trial record, Joseph Saikewicz was profoundly intellectually disabled, with a mental age of around two years and eight months. He had been diagnosed with leukemia at the age of 67, and the question arose whether to treat with an arduous course of chemotherapy that typically achieved a remission of two to thirteen months in thirty to fifty per cent of cases.

Although the court stated that it was employing a substituted judgment approach, it actually was employing a best interests approach. Indeed, inasmuch as Saikewicz had never been

competent, it made no sense to employ substituted judgment. The analysis of the *Saikewicz* court follows:

> . . . In short, the decision in cases such as this should be that which would be made by the incompetent person, if that person were competent, but taking into account the present and future incompetency of the individual as one of the factors which would necessarily enter into the decision-making process of the competent person. . . .
>
> The two factors considered by the probate judge to weigh in favor of administering chemotherapy were: (1) the fact that most people elect chemotherapy and (2) the chance of a longer life. . . . With regard to the second factor, the chance of a longer life carries the same weight for Saikewicz as for any other person, the value of life under the law having no relation to intelligence or social position. Intertwined with

> this consideration is the hope that a cure, temporary or permanent, will be discovered during the period of extra weeks or months potentially made available by chemotherapy. The guardian ad litem investigated this possibility and found no reason to hope for a dramatic breakthrough in the time frame relevant to the decision.
>
> The probate judge identified six factors weighing against administration of chemotherapy. Four of these—Saikewicz's age,[2] the probable side effects of treatment, the low chance of producing remission, and the certainty that treatment will cause immediate suffering—were clearly established by the medical testimony to be considerations that any individual would weigh carefully. A fifth factor—Saikewicz's inability to cooperate with

the treatment—introduces those considerations that are unique to this individual and which therefore are essential to the proper exercise of substituted judgment. The judge heard testimony that Saikewicz would have no comprehension of the reasons for the severe disruption of his formerly secure and stable environment occasioned by the chemotherapy. He therefore would experience fear without the understanding from which other patients draw

strength. The inability to anticipate and prepare for the severe side effects of the drugs leaves room only for confusion and disorientation. The possibility that such a naturally uncooperative patient would have to be physically restrained to allow the slow intravenous administration of drugs could only compound his pain and fear, as well as possibly jeopardize the ability of his body to withstand the toxic effects of the drugs.

The sixth factor identified by the judge as weighing against chemotherapy was "the quality of life possible for him even if the treatment does bring about remission." To the extent that this formulation equates the value of life with any measure of the quality of life, we firmly reject it. A reading of the entire record clearly reveals, however, the judge's concern that special care be taken to respect the dignity and worth of Saikewicz's life precisely because of his vulnerable position. The judge, as well as all the parties, was keenly aware that the supposed inability of Saikewicz, by virtue of his mental retardation, to appreciate or experience life had no place in the decision before them. Rather than reading the judge's formulation in a manner that demeans the value of the life of one who is mentally retarded, the vague, and perhaps ill-chosen, term "quality of

life" should be understood as a reference to the continuing state of pain and disorientation precipitated by the chemotherapy treatment. Viewing the term in this manner, together with the other factors properly considered by the judge, we are

334

satisfied that the decision to withhold treatment from Saikewicz was based on a regard for his actual interests and preferences and that the facts supported this decision. Superintendent of Belchertown State School v. Saikewicz, 370 N.E.2d 417, 431–432 (Mass. 1977).

The court's listing of factors seems reasonable, but the case for withholding chemotherapy was difficult given the court's concession that most competent people in Saikewicz's position would choose to be treated. The court therefore had to identify some significant factor weighing against treatment that was present here but not present generally. The real issue seemed to be that Saikewicz, perhaps on account of his incomprehension, would have to be restrained. Richard Burt, in an insightful commentary in Taking Care of Strangers (1979), concluded that in the trial court at least, the problems of administering the treatment were crucial to the decision against it. The trial transcript suggests that the judge was on the verge of ordering treatment when the attending physicians, present at the hearing, voiced their concern about being required to enter into a struggle with Saikewicz: "When you approach him in the hospital he flails at you and there is no way of communicating with him and he is quite strong; so he will have to be restrained."

Yet there was never any effort to determine whether Saikewicz's cooperation could somehow be obtained, perhaps with the aid of

institutional staff familiar to him, or by use of sedatives. Burt wrote

335

that "this omission reflected everyone's unwillingness to enter into sustained interaction with Joseph Saikewicz, everyone's wish to absent themselves from any transaction with him."

One cannot know for certain why the court did not examine the possibilities for administering treatment to Saikewicz more carefully before deciding to let him die, but the case provides a precautionary note about the risk that biases against the disabled can infect the assessment of best interests.

Continue treatment. In some states for some patients, courts and legislatures have decided that in the absence of clear and convincing evidence of the patient's wishes, treatment should be maintained. In the typical case, the patient is neither permanently unconscious nor terminally ill. The patient might have advanced Alzheimer's disease and a life expectancy of two years or more. In other cases, the patient has suffered serious and irreversible neurologic injury. If these patients did not clearly indicate a desire not to be treated, the courts will require that treatment be provided.

For a striking example of a court requiring the provision of treatment in the absence of clear and convincing evidence that the patient would refuse it, consider the Michael Martin case. According to the Michigan Supreme Court, for patients who are neither terminally ill nor permanently unconscious and do not have a written advance directive, prior oral statements will be sufficient to justify withdrawal of treatment "[o]nly when the patient's

336

prior statements clearly illustrate a serious, well thought out, consistent decision to refuse treatment under these exact circumstances or circumstances highly similar to the current situation." In re Martin, 538 N.W.2d 399, 411 (Mich. 1995).

In Martin's case, this test was not satisfied despite considerable testimony suggesting he would not want treatment provided. He had suffered serious injuries, including a head injury, from an automobile accident that left him able to understand "only very short and very simple questions." He could not "accurately comprehend questions that are lengthy, verbose, or that require the retention of multiple thoughts." Martin's wife testified to eight years of discussions regarding his wishes in the event of a serious accident or disabling illness, the most recent discussion occurring one month before his accident. These discussions took place after the Martins watched movies about people who could no longer take care of themselves because of an accident or illness.

> Mike stated to me on several occasions: "That's bullshit. I would never want to live like that." He also said to me, "Please don't ever let me exist that way because those people don't even have their dignity." . . . Mike was an avid hunter and frequently expressed concerned [sic] about a hunting accident. Mike frequently told me that if he ever had an accident from which he would "not recover" and "could not be the same person," he did "not want to live that way." He would say, "Mary, promise me you wouldn't let me live like

> that if I can't be the person I am right now, because if you do, believe me I'll haunt you every day of your life." Id. at 412.

Often, in states that call for continued treatment in the absence of clear and convincing evidence, if treatment would be extraordinarily burdensome or inhumane, family members could refuse treatment. The *Conroy* case is illustrative. Claire Conroy was an elderly nursing home patient; she was not unconscious, but suffered from severe, permanent, mental and physical impairments —very much like someone with advanced Alzheimer's disease. She had left no clear instructions. Her life expectancy was no more than about one year regardless of what was done for her, but death was not imminent.

What does *Conroy* hold? If clear and convincing evidence about the patient's wishes is lacking, doctors can turn to two more objective standards. First, where there is some "trustworthy" evidence that the patient would have refused further treatment (but that evidence falls short of meeting the "clear and convincing" test), treatment may be withdrawn if the burdens of the patient's continued life with treatment "markedly outweigh" the benefits the patient may derive from that life. According to the court, this means that the patient "is suffering, and will continue to suffer throughout the expected duration of his life, unavoidable pain." 486 A.2d at 1232.

Conroy's second objective test applies where there is no trustworthy evidence of the patient's preference. In such a case, it is not enough that the

pain and suffering of continued life "markedly outweigh" the benefits. Before withdrawal of treatment can be allowed for such a patient, one must also find that the patient will suffer so much pain that prolonging life "would be inhumane." Id.

The *Conroy* majority's emphasis on pain reflects a common feature of end-of-life law in the United States—that there be an objective measure of patient suffering. This desire for an objective measure also appears in the concurring opinions of Justices Stevens and O'Connor in the *Glucksberg* aid-in-dying case, as well as in the common requirement of death with dignity acts that the patient be terminally ill.

c. General Trends in Procedural Rules

Despite the state-to-state variation, some patterns do emerge for patients who have not left clear and convincing evidence of their wishes. Family decision-making is a common approach, both through statutes and court decisions.

Perhaps even more common is for states to adopt stricter procedural rules when the incompetent patient is neither terminally ill nor permanently unconscious, as in New Jersey with the *Conroy* case and Michigan and the *Martin* case. In such cases, the patient has some degree of consciousness and may live for a year or more. Absent clear and convincing evidence that the patient would not want treatment, these states require treatment or allow for the treatment to be withheld if the patient would seriously suffer from continuing with treatment. On the other hand, for patients who are terminally ill or

permanently unconscious, as in *Jobes*, family decision-making is common.

Note the implications of varying the procedural rules depending on the patient's medical condition. States such as New Jersey and Michigan have effectively made the patient's prognosis a factor in

whether treatment may be withdrawn from a patient who has lost decision-making capacity. As we saw earlier, the competent patient's right to refuse life-sustaining treatment generally does not depend on whether the patient can live for 50 minutes or 50 years; nor does it depend on whether the patient can be restored to perfect health or be left with serious disabilities. For the incompetent patient, on the other hand, it is easier to justify withdrawal of treatment when the patient is either permanently unconscious or terminally ill. If a state has different procedural standards for patients like Claire Conroy or Michael Martin than for patients like Karen Quinlan or Nancy Cruzan, then it no longer is true that the right to refuse life-sustaining treatment is not altered by incompetence. As a practical matter, stricter procedural rules make for a weakening of the right. And the impact is likely to be greater for patients who have a lower socioeconomic status, who are less likely to know about advance directives.

Other state supreme courts have followed this approach, including California (in Conservatorship of Wendland, 28 P.3d 151 (Cal. 2001)), and Wisconsin (in Spahn v. Eisenberg, 563 N.W.2d 485 (Wis. 1997)). New York has taken this approach by statute.

340

Although it is controversial whether clear and convincing evidence of the patient's wishes should be required before discontinuing life-sustaining treatment, it is less controversial when courts impose strict standards for establishing the patient's prognosis. If life-sustaining treatment is withdrawn because the patient is terminally ill or permanently unconscious, for example, we would want to be quite sure that the patient really is terminally ill or permanently unconscious. The Alabama Supreme Court came to that conclusion when it adopted a clear and convincing evidence

standard for deciding that a patient is permanently unconscious for purposes of carrying out the patient's advance directive in the event of permanent unconsciousness. Knight v. Beverly Health Care Bay Manor Health Care Center, 820 So. 2d 92, 102 (Ala. 2001).

d. Advance Directives

As court opinions regularly observe, people can avoid many of the problems with end-of-life decision-making by executing a living will, durable power of attorney for health care, or other advance directive while competent. All states have either a living will or power of attorney statute and almost all states have both.

The statutes typically state that the rights granted therein are cumulative with other rights individuals might have with respect to end-of-life medical decisions. Thus, if a patient prepares a written directive that doesn't comply with the requirements of the state's living will statute, the directive still

would have legal force under the common law and constitutional rights to have treatment withdrawn on the basis of clear and convincing evidence of the patient's wishes.

Living Wills. The first living will statute was enacted in California in 1976; the document is called a "living" will because it takes effect while the testator is still alive. With a living will, also called a treatment directive, a person describes the circumstances under which treatment would not be desired. A person also might use a living will to indicate a desire that treatment be provided as long as life can be prolonged.

Living wills have several drawbacks. If a person gives specific instructions, the document will not provide guidance in

unanticipated situations. If a person instead tries to give general guidance that can be applied to any particular situation, then there will likely be a good deal of ambiguity in the living will, ambiguity that may lead different people to different interpretations of the person's wishes. Some statutes apply only to patients with terminal illnesses and persistent vegetative states, and some statutes restrict the circumstances under which artificial nutrition and hydration may be withdrawn. People can avoid such limitations by attaching addenda or using one of the model forms available from national organizations, although most people are likely to use the statutory form out of either convenience or unawareness of the alternatives.

Proxy Appointments. Rather than giving treatment instructions, a person might choose to

appoint a proxy or surrogate decision-maker by executing a durable power of attorney for health care. By doing so, the individual can transfer authority to make medical decisions to someone else.

Durable powers of attorney were created because under the common law, all agency power ceases when the principal becomes incompetent. With a "durable" power of attorney, the proxy's agency authority survives the incompetence of the patient. Technically, durable powers of attorney for health care are actually "springing" powers of attorney. In other words, strictly speaking, a durable power of attorney takes effect while the principal is still competent and continues to have effect if the principal becomes incompetent. A springing power of attorney, on the other hand, does not take effect until the principal becomes incompetent. Despite the inaccuracy of the term, powers of attorney for health care are universally characterized as durable powers of attorney.

With a power of attorney, a person can avoid many of the problems with living wills. Power of attorney statutes generally apply to patients in any condition and with regard to any treatment. It is unusual for them to qualify the surrogate's authority to situations in which the patient is terminally ill or to restrict the surrogate's authority to decline artificial nutrition and hydration. Accordingly, people can provide for the exercise of their right to refuse treatment in all circumstances. Since proxies have broad authority, they are able to make decisions even when patients have not expressed their wishes or

expressed them in too vague a way to be sure what the patient intended. Of course, some people may not have anyone whom they trust enough to appoint as a proxy.

Some people choose to combine a treatment directive with a power of attorney, giving some instructions and leaving authority to the surrogate for situations not covered by their instructions. The problem with this approach is that family members or health care providers who disagree with the surrogate can argue that the surrogate is not following the patient's instructions in the treatment directive. Accordingly, it may be better to give instructions to the surrogate privately.

Recall that when patients do not leave a living will to express their wishes about the use of a ventilator, feeding tube, or other treatment, courts will consider oral statements by patients to see whether there is clear and convincing evidence of the patients' treatment preferences. If health care providers should carry out patient wishes for a treatment decision when patients have expressed their preferences through clear oral statements rather

than through a formal written document, they also should consider oral statements by patients regarding their choice of surrogate. For example, it was quite clear that Michael Martin expected his wife to make medical decisions for him. This question has not been addressed by courts, so it is typically assumed that surrogates need to be authorized by a written appointment or a surrogate decision-making statute.

344

Statutory Forms. Advance directive laws typically include a statutory form but generally state that individuals need not use the statutory form. However, in a few states, the statutes require compliance with the statutory form. But even in these states, a handwritten living will should be valid. As mentioned, since treatment always can be withheld or withdrawn if the patient has left clear and convincing evidence of a desire not to be treated, written documentation in any form should be effective.

Limitations of Advance Directives. For many years, advance directives were disappointing in practice. Most people did not fill them out, even when encouraged to do so by physicians. Moreover, as indicated, when living wills were executed, they often were too vague to give sufficient guidance. And even when living wills gave sufficient guidance, they often were overridden by physicians.

Thus, for example, while courts clearly recognized that end-of-life decisions should be based on patient preferences and values, empirical studies indicated that the physician's preferences and values seemed to drive decisions regarding the withdrawal of life-sustaining medical treatment. This predominance of physicians' values also could be found in situations involving competent

patients or in which family members had decision-making authority.

In recent years, there has been improvement. Dying patients are much more likely to have completed an advance directive, and studies have found closer agreement between patient preferences

and the care they received. But the data are mixed on agreement between patient preferences and care received, and failures to respect patient preferences seem to be more likely when patients request aggressive care.

e. Deciding for Children

(1) The General Framework of Child Protective Laws

Legally, the child's situation is different from that of the incapacitated adult. As we have seen, states vary in the extent to which they recognize family decision-making for formerly competent adults. But in the case of a child, the parents acquire decision-making authority at the child's birth. As the Supreme Court has recognized, parents have a fundamental right to make decisions about the upbringing of their children, both because they are most concerned about the well-being of their children and also to avoid undue efforts by the state to control the destiny of its citizens.

But as with other rights, parental rights are subject to reasonable limits. For purposes of this chapter, the state "can intervene in the parent-child relationship where the health and safety of the child . . . are in jeopardy." Newmark v. Williams, 588 A.2d 1108, 1116 (Del. 1991). Thus, under their child abuse and neglect laws, states

provide for legal authority to override parental decisions that would constitute medical neglect.

346

The classic cases are those involving families of Jehovah's Witnesses where the children need blood transfusions and their parents decline the transfusions on religious grounds. In those cases, the courts have held that the children must be given the transfusions. In ordering treatment, the courts typically cite Prince v. Massachusetts, where the Court wrote:

> Parents may be free to become martyrs themselves. But it does not follow they are free, in identical circumstances, to make martyrs of their children before they have reached the age of full and legal discretion when they can make that choice for themselves. 321 U.S. 158, 170 (1944).

The hard question in this area is where the limits of parental authority end, and when a refusal of treatment becomes child neglect or abuse. At either end of the spectrum, the law is fairly clear.

Courts have had no trouble holding that, when a child can readily be restored to good health or when treatment poses little risk, parents may not refuse a life-sustaining treatment, such as a blood transfusion, antibiotics, or an appendectomy. Similarly, courts have had no trouble holding that parents may refuse life-sustaining ventilators or feeding tubes when the child is irreversibly unconscious.

Other decisions, however, are not so clear. Some treatments will have a low likelihood of success and may carry substantial risks. As the case law

indicates, the courts have not come to a consensus on the extent of parental discretion. Consider, for example, the following two cases:

In one of the cases, a three-year-old boy was diagnosed as having an aggressive malignancy of the immune system, Burkitt's lymphoma. By the time a diagnosis was made during surgery for an intestinal obstruction, the cancer had spread elsewhere in the boy's body. His physician recommended that he be treated with a heavy regimen of chemotherapy. According to the medical testimony, there was a 40 percent chance of a "cure." Without treatment, he would likely die within six to eight months. Medical testimony also indicated that the treatment itself was quite toxic and might prove fatal. The boy's parents declined the chemotherapy in favor of care from a Christian Science practitioner. Given the substantial risks of treatment and the less than 50 percent chance of success, the court upheld the pa-rental refusal of treatment. *Newmark, supra*.

In the other case, a 12-year-old girl was diagnosed with Ewing's sarcoma, a bone cancer for which the girl had a 25 to 50 percent chance of long-term remission with treatment. Without treatment, she would likely die within six to nine months. Her father, a lay minister of the Church of God of the Union Assembly, refused treatment on religious grounds. Given the apparent certainty of death without treatment, and the reasonable possibility of long-term success with treatment, the court overrode the parental refusal of treatment. The court seemed to consider the fact that the family's religious sect did

not refuse all kinds of medical treatment. In re Hamilton, 657 S.W.2d 425 (Tenn. Ct. App. 1983).

In other illustrative cases, courts have permitted a mother to refuse aggressive antiviral drug therapy to treat HIV infection in her four-year-old son, and deferred to a parent's refusal of antipsychotic drugs for three-year-old child, but denied the parents' request to withdraw a ventilator from a 14-year-old child who had a lethal genetic disease and a life expectancy of no more than two years but who was still "aware and enjoy[ed] TV and videos." In other cases, the courts imposed treatment for cancers that can be quite responsive to treatment. Hall, Orentlicher, Bobinski et al., *supra*, at 577.

Many states provide exemptions in their child abuse and neglect laws for parents who refuse medical treatment on religious grounds. In some of these states, the exemptions excuse parents from liability when their children suffer from the denial of care, though generally not when the child's life is endangered. In addition, the exemptions typically apply to child and abuse laws but not so often to other laws under which parents may be held accountable, such as involuntary manslaughter statutes. And of course, the exemptions speak to the ability of the state to punish parents. Even if a court cannot sanction the parents, it still may order that treatment be provided to the child. Indeed, the federal regulation that led to the adoption of the religious exemptions explicitly drew a line between holding parents liable and ensuring access to care for children. The regulation, which no longer is in effect,

tied federal funding for child abuse programs to the enactment of religious exemptions. But the regulation also stated that the exemptions "shall not preclude a court from ordering that medical services be provided to the child, where his health requires it." State law exemptions typically track this requirement. In practice, though, the judicial power to order treatment does not always provide adequate protection for children.

(2) The Special Case of the Newborn

Treatment decisions for severely disabled newborns can be particularly challenging because it is more difficult to assess prognosis in infants than older persons. The child's ultimate disabilities may not be as severe as predicted at birth.

These decisions generated considerable debate in the 1980s in conjunction with several prominent cases in which parents sought to withhold life-sustaining treatment from children born with Down syndrome or other, more serious conditions. One case involved the death in 1982 of a Bloomington, Indiana, infant with Down syndrome. The child suffered from several congenital abnormalities associated with Down syndrome, including an esophageal obstruction that prevented oral feeding but that could have been corrected by surgery. The parents declined surgery, and the child died after a trial court and the local Child Protection Committee upheld the parents' decision. In response to the case, the U.S. Department of Health and Human Services adopted rules under the Rehabilitation Act of 1973 to regulate

treatment decisions for severely disabled newborns, but the Supreme Court invalidated the rules as not justified by any evidence that such persons were being discriminated against in the

provision of health care. Bowen v. American Hospital Association, 476 U.S. 610 (1986).

(a) Child Abuse Amendments of 1984

Also in response to the controversy generated by the Bloomington case, Congress enacted provisions of the Child Abuse Amendments of 1984 to address concerns about the "withholding of medically indicated treatment from infants with disabilities who have life-threatening conditions." 42 U.S.C. § 5106a(b)(2)(C)).

The amendments condition state receipt of certain federal funds on assurance that the state has in place procedures for responding, under state child abuse laws, to cases of "medical neglect, which include the "withholding of medically indicated treatment." "Medically indicated treatment" is then defined as treatment which, in the treating physicians' "reasonable medical judgment, will be most likely to be effective in ameliorating or correcting all [of the infant's life-threatening] conditions, except that the term does not include the failure to provide treatment . . . to an infant when, in the treating . . . physicians' reasonable medical judgment:

(A) the infant is chronically and irreversibly comatose;

(B) the provision of such treatment would (i) merely prolong dying, (ii) not be effective in ameliorating or correcting all of the infant's life threatening conditions, or (iii) otherwise be futile in terms of the survival of the infant; or

(C) the provision of such treatment would be virtually futile in terms of the survival of the infant and the treatment itself under such circumstances would be inhumane.

42 U.S.C. § 5106g(a)(5).

These federal standards are relatively extreme. In general, they appear to require continued treatment in all cases in which life may thereby be extended; the only clear exception is the case of a newborn who is irreversibly unconscious. What then of a child, such as those with the most severe cases of spina bifida, for whom life may be extended, but at a cost in pain and discomfort that many would feel is too great to pay for the little pleasure the child will be able to derive from the added years? The last provision, subsection (C), provides the only potential flexibility for allowing discontinuance of treatment in such a case. But that provision seems unlikely to apply, for we probably could not conclude that the treatment would be "virtually futile in terms of the survival of the infant."

As a practical matter, the reach of the federal statute is limited because it is a condition for grants, not a substantive standard directly applicable to parents, physicians, or hospitals. Moreover, while there have been cases in which the law was at issue,

the courts generally have concluded that the parents could proceed with their decision to decline treatment.

However, a decision by a Wisconsin court of appeals took a strict view of the Child Abuse Amendments. In a case involving a child born after 23 weeks of gestation who was given maximal life-prolonging treatment, the parents sued, claiming that they were not sufficiently informed of their child's prognosis when they agreed to some of the life-saving measures. The court rejected the parents' claim, concluding that under both the Amendments and Wisconsin end-of-life law, the parents did not have the right to refuse life-

sustaining treatment for their infant. Montalvo v. Borkovec, 647 N.W.2d 413 (Wis. Ct. App. 2002).

Even if courts defer to parental decisions, parental wishes may be frustrated by a perception among physicians that the federal law mandates care. Early data suggested that the statute was having a significant impact on physician practices, but more recent reports suggest that there may not have been a substantial impact overall from the Child Abuse Amendments on clinical practice. Charity Scott, *Baby Doe at Twenty-Five*, 25 Ga. St. U. L. Rev. 801, 807–811 (2009).

(b) Rehabilitation Act and Americans with Disabilities Act

While there has not been much litigation over the Child Abuse Amendments of 1984, there has been some important litigation of treatment decisions for

severely disabled newborns under § 504 of the Rehabilitation Act of 1973. Section 504 prohibits recipients of federal funds from discriminating against "otherwise qualified" disabled persons solely on the basis of their disability. 29 U.S.C. § 794. The Americans with Disabilities Act of 1990 extends the protections of § 504 to virtually all health care settings.

In the § 504 cases, the claim is that denying treatment to the newborn is an unlawful act of discrimination. In these cases, the courts have permitted the withholding of treatment. For example, in United States v. University Hospital, 729 F.2d 144 (2d Cir. 1984), the issue was whether surgery could be withheld from an infant with spina bifida, a condition of variable severity in which there is incomplete closure of the tissues surrounding the spinal cord. In

this case, the infant had serious physical problems like impaired bowel and bladder function and was thought to have such compromised mental function that she would "never interact with her environment or other people." Id. at 146. The surgery, which would have closed the opening in her spine and implanted a shunt to drain excessive fluid buildup in her brain, was likely to prolong the infant's life but not expected to do anything to treat her disabilities. A state appellate court had refused to intervene, noting that, while the surgery would enhance the infant's chances of living, it also might aggravate her disabilities. In also refusing to intervene, the Second Circuit wrote:

> Doe [v. New York University, 666 F.2d 761 (2d Cir. 1981),] establishes that § 504 prohibits discrimination against a handicapped individual only where the individual's handicap is unrelated to, and thus improper to consideration of, the services in question. As defendants here point out, however, where medical treatment is at issue, it is typically the handicap itself that give rise to, or at least contributes to, the need for services. Defendants thus argue, and with some force, that the "otherwise qualified" criterion of § 504 cannot be meaningfully applied to a medical treatment decision. Similarly, defendants argue that it would be pointless to inquire whether a patient who was affected by a medical treatment decision, was "solely by reason of his handicap . . . subjected to discrimination."
>
> . . . Where the handicapping condition is related to the condition(s) to be treated, it will rarely, if ever, be possible to say with certainty that a particular decision was "discriminatory." 729 F.2d at 156–157.

See also Johnson v. Thompson, 971 F.2d 1487 (10th Cir. 1992) (finding no violation of § 504 when surgery was withheld from children with spina bifida).

In the end, the parents in University Hospital agreed to have a shunt implanted to drain the fluid in their daughter's brain, although the surgery was delayed because of an infection that was likely related to the opening in her spine. The child, Keri-Lynn, has done much better than predicted.

Although she is confined to a wheelchair and developmentally disabled, she can talk, and by age 30 in 2013, she was living in a group home Monday through Friday taking academic classes and receiving physical therapy. It is not clear whether surgery to close Keri-Lynn's spine would have improved her outcome.

(c) The Role of Physicians

There have been important changes with respect to parental discretion in recent decades because of changes in physicians' views as to what constitutes appropriate care for seriously disabled newborns. Partly because advances in medical care are resulting in better outcomes and partly because of changing social attitudes about disabled infants, physicians are becoming more aggressive in recommending treatment. For example, while many physicians once believed it reasonable to withhold surgery to correct an intestinal obstruction in a child with Down syndrome and allow the child to die, it would probably be very difficult today to find a physician taking that position. Since courts are likely to defer to physicians as to whether treatment is necessary, the range of parental discretion has been narrowing.

B. PHYSICIAN AID IN DYING

We have seen that the competent patient has a broad right to refuse medical treatment even though death might result. An important question is whether this strong right should be extended to

permit a patient to hasten death by taking a lethal dose of a drug.

As a preliminary matter, it is worth considering the different descriptive terms used for this topic. Although courts commonly talk about "physician-assisted suicide," others prefer "physician aid in dying," "death with dignity," or similar terms. Choice of terms depends a good deal on whether one views decisions by dying persons to manage their deaths with self-administered medications as "suicides" or more like decisions to refuse life-sustaining medical treatment. Following the recommendation of the American Public Health Association, this text will use the term "physician aid in dying."

In some countries, aid in dying includes both patient administration of the lethal drug and physician (or nurse) administration. In the United States, only patient administration is permitted.

1. AID IN DYING AND THE CONSTITUTION

Seven years after the *Cruzan* case, the U.S. Supreme Court took up the question of physician aid in dying. In the *Glucksberg* case in 1997, the Court rejected a right to aid in dying, observing that the right in *Cruzan* was not just a matter of personal autonomy. Rather, it rested on a "long legal tradition protecting the decision to refuse unwanted medical treatment." In contrast, wrote the Court, the

"decision to commit suicide with the assistance of another . . . has never enjoyed similar legal protection." Washington v. Glucksberg, 521 U.S. 702 (1997).

357

In *Glucksberg*, the Court considered a substantive due process challenge to a Washington law that made it a felony punishable by up to five years' imprisonment to "knowingly cause or aid another person to attempt suicide". The challenge was brought by physicians who complained that the law kept them from providing their terminally ill patients the assistance they sometimes wanted to end their lives. The physicians were joined by three seriously ill patients and Compassion in Dying. If the Court had struck down such provisions on their face, the effect would have been to hold that the Constitution compels every state to allow what Oregon, Washington, and several other states have now done by statute.

Framing the question before it narrowly—whether the liberty protected by the Due Process Clause includes "a right to commit suicide which itself includes a right to assistance in doing so"—the Court answered "no". Five justices signed the opinion for the Court, but all nine concurred in the judgment. Agreeing that many of the rights found within protected liberty "sound in personal autonomy", the Court nonetheless rejected the "sweeping conclusion" that all important intimate decisions are so protected, and relied upon the law's historically consistent criminalization of assisted suicide to conclude that it was outside the Due Process Clause's protected zone.

It is not surprising that the Court was more reluctant to recognize a right to aid in dying than a right to refuse treatment. Most people see greater

risks to patient welfare from a right to aid in dying. In addition, by the time the Court decided the *Cruzan* case, there had been nearly a decade-and-a-half of experience with refusals of life-sustaining treatment since the *Quinlan* case, and both the medical profession and the public were comfortable with end-of-life decision-making. When the Court decided *Glucksberg*, there had not yet been any experience with aid in dying in the United States, and there were significant concerns about the experience with aid in dying in the Netherlands.

2. RIGHTS TO AID IN DYING

Although the Supreme Court did not recognize a constitutional right, several states have enacted legislation permitting aid in dying. Starting with Oregon in 1997, a right to aid in dying has gradually spread to other states, including California, Colorado, and New Jersey. In Montana, the state supreme court held that aid in dying is not prohibited by any law.

A key difference between the right to aid in dying and the right to refuse treatment is that the aid in dying right does not include all patients, regardless of their medical condition. Rather, only patients who are terminally ill—whose life expectancy is six months or less —can choose aid in dying.

There are a few other important limits on the aid in dying right. As mentioned above, the patient must self-administer the drug. In addition, patients must be competent adults who confirm their desire for aid in dying over a 15-day or longer period of time.

Accordingly, the right to aid in dying does not survive incompetence.

With these restrictions on the right, many of the concerns about aid in dying have not materialized. Oregon has the longest experience with legalized aid in dying—more than twenty years—and less than one percent of deaths in Oregon are by that practice. The most common diagnosis has been cancer, and the patients have been similar to other dying patients with regard to sex, race, urban or rural residence, health insurance coverage, and hospice enrollment. The level of education has been higher for patients who choose aid in dying. The most common concerns of patients choosing aid in dying have been loss of autonomy, decreased ability to participate in enjoyable activities, and loss of dignity.

3. THE DISTINCTION BETWEEN TREATMENT WITHDRAWAL AND AID IN DYING

When the Supreme Court rejected a right to aid in dying in *Glucksberg* and the companion case, Vacco v. Quill, 521 U.S. 793 (1997), the Court invoked a number of arguments that are commonly made to distinguish the right to refuse life-sustaining treatment from a right to aid in dying. As we will see, while the Court raised a number of important concerns, it's not clear that their arguments really distinguish aid in dying from the withdrawal of life-sustaining treatment for a patient who is terminally ill.

Causation and the Natural Death. For example, it often is said that when life-sustaining

treatment is withdrawn, patients die from the natural progression of their disease rather than being killed by a lethal drug. *Vacco*, 521

U.S. at 801. This argument makes two common distinctions—first, between deaths allowed to happen and deaths affirmatively caused, and second, between "natural" and "unnatural" deaths.

With regard to causation, the doctor who turns off a ventilator causes the patient's death just as does the doctor who writes a prescription for a lethal dose of medication. Indeed, if a doctor were to turn off all of the ventilators of patients receiving intensive care, the physician would be charged with murder, and it would be no defense that the physician let the patients' diseases take their natural course.

As to the rejection of unnatural deaths, it is not clear why we should prefer natural deaths over unnatural ones. Virtually all medical treatments—including joint replacements, heart surgery, and kidney dialysis—are unnatural. As these examples illustrate, often the unnatural is better than the natural. From the perspective of the terminally ill patient, the "unnatural" death of aid in dying is preferable to a "natural" death because it relieves the patient of unbearable suffering.

What this suggests is that self-killings are not automatically wrong, Rather, as one philosophical tradition has it, suicide refers only to *wrongful* self-killings, just as murder is the *wrongful* killing of another. The real question then lies in deciding when a self-killing is wrongful. Killing oneself to avoid pain or unhappiness would usually be regarded as suicide,

but there may be situations in which a self-killing would be justifiable, as in the case of a terminally ill patient possessing decision-making capacity.

Physician Intent. When a physician writes a prescription for a lethal medication, it is said, the physician necessarily intends the patient's death. *Vacco*, 521 U.S. at 802. A physician who withdraws life-sustaining treatment can intend only to relieve the burdens of treatment and hope that the patient survives the withdrawal.

There are two important responses to this argument. First, the law generally holds people accountable for the foreseeable consequences of their actions, even when unintended. The company that maintains a workplace dangerous to employee safety can be punished if a worker dies on the job. In addition, doctors who write a prescription for a lethal drug may hope that the patient never takes the medication or doesn't even fill the prescription. In Oregon, only about 65 percent of patients who receive an aid-in-dying prescription die from taking the pills. Physicians can genuinely hope that their patients will be among the 35 percent who do not ingest the medication. For these patients, knowing that they can turn to aid in dying provides an important source of reassurance during a very difficult time.

"Negative" Versus "Positive" Rights. In the United States, individuals enjoy negative rights to be free from government interference, but not positive rights to assistance. In this view, the negative right to refuse unwanted treatment does not imply a positive right of access to lethal medication.

However, a right to aid in dying also is a negative right. Having that right ensures that the government will not interfere when a physician decides to write a prescription for the aid-in-dying medication.

The positive right would be a right of patients to insist that their physicians provide aid-in-dying prescriptions, but all aid-in-dying laws make clear that physicians are not obligated to participate in the practice. Abortion provides a useful analogy. Under the negative right to abortion, the government cannot interfere when a physician is willing to perform an abortion, but there is no positive right to abortion that would require physicians to perform abortions when requested by their patients.

Trust in Physicians. If doctors began to dispense death-causing agents, it is feared, patients would develop a profound distrust of the medical profession. Physicians are supposed to heal patients, not end their lives. *Glucksberg*, 521 U.S. at 731.

But physicians also are supposed to relieve suffering and ease discomfort. As we saw with the medical profession and the right to refuse treatment, at some point, treatment turns from trying to cure the curable to trying to ensure a comfortable dying process for the incurable. For some patients, a comfortable dying process means aid in dying.

At any rate, if physicians were supposed to only try to heal, it would not be permissible to withdraw life-sustaining treatment. That too entails an act to shorten rather than prolong a patient's life.

And there is no reason to think that physician participation in treatment withdrawals or aid in dying has led to a decline in patient trust in the medical profession. Nor has physician participation in abortion seemed to have led to a decline in patient trust.

Risks of Abuse. For a number of reasons, many people worry that legalizing aid in dying will lead to misuse of the practice. Terminally ill patients might be pressured to choose aid in dying

because they have become a burden on their families or the health care system. *Glucksberg*, 521 U.S. at 731–732. Requests for aid in dying may be made by depressed patients, yet many physicians fail to recognize treatable depression in their patients, particularly the elderly. Id. at 730. In addition, many patients receive inadequate palliative care, and may choose aid in dying when good palliative care would address their suffering.

All of these concerns are important, but they do not distinguish aid in dying from the withdrawal of life-sustaining treatment. Patients may refuse a ventilator, dialysis, or other therapy because they are depressed, because they have not received adequate palliative care, or because they have been pressured to do so by family members or health care providers. Indeed, when it comes to concerns about the costs of care, withdrawals of treatment from patients who could live years, even decades, can save much more money than can aid in dying for patients with a life expectancy of six months or less.

For the withdrawal of treatment, the risk of abuse had led to the adoption of safeguards to protect against the risk, not a prohibition against withdrawal. For example, when the New Jersey Supreme Court recognized family decision-making in the *Jobes* case, it wrote that if "a health-care professional becomes uncertain about whether family members are properly protecting a patient's interests, termination of life-sustaining treatment should not occur without the appointment of a guardian." 529 A.2d at 447. Similarly, the law can guard against the risk of abuse with aid in dying by adopting safeguards, as all legalizing states have with their requirements of terminal illness, decision-making capacity, waiting periods, self-administration, etc.

Indeed, the safeguards adopted for aid in dying in the United States directly address concerns about aid in dying practices in the Netherlands and Belgium. The controversies over particular aid-in-dying cases in the Netherlands and Belgium, as well as over particular illicit aid-in-dying cases in the United States, typically involve patients who were not terminally ill or who had diminished mental capacity. For example, a patient featured in Frontline's "The Suicide Plan" harbored false beliefs about her health because of mental illness, and a physician in the Netherlands was prosecuted and acquitted on charges after he administered a lethal

injection to a patient with dementia on the basis of an earlier request for aid in dying.[3]

Making Sense of the Distinction. If the usual arguments for the distinction between withdrawal of treatment and aid in dying don't seem to explain the distinction, how do we explain it?

Consider the possibility that aid in dying was prohibited for the same reasons that the right to refuse treatment evolved from a right only for the seriously ill in *Quinlan* to a right for all persons.

As we saw, courts may have been uncomfortable with the idea of the government deciding which lives must be prolonged and which lives need not be sustained. Accordingly, courts generally recognized a right to refuse treatment for all person, knowing that refusals would be asserted primarily by the seriously ill. But a lethal dose of medication might be taken by people who are not seriously ill. It would be too risky to allow aid in dying for all patients.

On the other hand, a total ban on aid in dying left an important gap in end-of-life law. There are many patients who are suffering

greatly from serious and irreversible illness and who are not dependent on a life-sustaining treatment. A right to aid in dying would allow these patients the same ability as patients dependent on a ventilator or feeding tube to avoid the further prolongation of the dying process. The law can meet the needs of these patients without

opening up aid in dying to other patients by limiting the aid in dying right to patients who are terminally ill.

4. VOLUNTARY STOPPING EATING AND DRINKING (VSED)

Some commentators have argued that patients who are not dependent upon life-sustaining treatment but who desire aid in dying can always end their lives by refraining from eating and drinking. For some patients, that is an important option.

But there are disadvantages to this alternative. VSED usually takes one-to-two weeks, prolonging the dying process for patients who are trying to end their suffering quickly. In addition, patients become very thirsty once they've stopped eating and drinking, exacerbating their discomfort. To some extent, these concerns are alleviated by the fact that many patients become very sleepy or even unconscious after a few days without food and water. In addition, there is some uncertainty about the legal status of a right to refrain from eating and drinking. Recall in this regard how Justice Sandra Day O'Connor's concurring opinion in *Cruzan* analogized artificial nutrition and hydration to other medical treatments.

5. PALLIATIVE SEDATION

In her concurring opinion in *Glucksberg*, Justice O'Connor suggested that dying patients have a constitutional right to alleviate their suffering, but that such a right would not imply a right to aid in

dying. As she observed, patients are able to obtain medications to relieve their suffering "even to the point of causing unconsciousness and hastening death."

It is well accepted in ethics and law that physicians can administer medications to relieve pain or other suffering even if doing so poses some increased risk of death for the patient. Under this principle of "double effect," the risk of death is acceptable as long as the medication is used in a reasonable effort to treat the patient's suffering. By way of analogy, we permit physicians to perform open heart surgery, de-spite the risk of patient death, because the primary purpose of the surgery is to treat the patient's heart disease, and the benefits of the surgery are reasonable when compared with the risks.

In some cases, as Justice O'Connor noted, the patient must be sedated into unconsciousness to relieve the suffering. In these cases, the patient is sedated into a coma from which the patient usually dies in a few days, either because the patient is at the end stage of the underlying illness or because food and water are withheld.

Note how the line blurs between palliative sedation and euthanasia. If a patient is sedated and artificial feeding withheld, the patient might not die from the underlying disease but from the combination of sedation and withholding of food and water. In other words, the patient's death would be caused by the

combination of two kinds of physician action rather than physician inaction.

368

States have begun to recognize a right to palliative care in their legislatures and courts. Florida, California, New York, and other states have passed statutes requiring physicians to advise terminally ill patients of their options for end-of-life care, including hospice care, and treatment to relieve pain and other symptoms. In addition to these statutory rights, patients may be able to recover damages from doctors and other health care providers for the failure to provide adequate pain relief.

If there is a right to receive medications to alleviate the symptoms of disease, is there also a right to receive medications to treat the disease itself? In a surprising decision in favor of patients' rights based on *Cruzan* and *Glucksberg*, a panel of the U.S. Court of Appeals for the D.C. Circuit initially found a limited constitutional right of terminally ill persons to receive experimental cancer chemotherapy or other potentially therapeutic drugs without Food and Drug Administration (FDA) approval when there are no other therapeutic options left. Abigail Alliance v. Eschenbach, 445 F.3d 470 (D.C. Cir. 2006). This decision was reversed by the *en banc* court, however, 495 F.3d 695 (D.C. Cir. 2007), which reasoned that such a right is not deeply rooted in the nation's history and tradition, considering the long history of the regulation of pharmacists and pharmaceuticals. As a result of the litigation, the FDA revised its regulations to give terminally ill patients greater access to experimental drugs.

369

C. DENYING "FUTILE" TREATMENT TO THE PATIENT WHO REQUESTS IT

If the principle of patient autonomy supports the withdrawal of care at the patient's request, what of the patient who *seeks* care that the physician or hospital is otherwise disinclined to provide?

While cases about life-sustaining medical treatment have generally involved situations in which the patient or patient's proxy wanted to stop treatment over the objection of physicians or the hospital, cases are also arising in which the positions are reversed: the patient or proxy wants to continue with treatment, but the physicians or hospital want to stop providing care. In such cases, the patient or family asserts the patient's right to make medical decisions; the physicians or hospital argue that the treatment is medically "futile," that it does not provide sufficient—or any—medical benefit and therefore ought not to be offered to the patient.

The increase in futility cases reflects a few factors. For example, if the right to refuse treatment reflects a sense that medical care at the end of life can cause harm rather than deliver benefit, physicians may become uncomfortable providing the care. Also, the advances in medical technology that allow doctors to prolong life can be very expensive and stretch the limited resources of the state or private insurers. If we cannot afford to provide all care for all patients, it makes sense to cut back on care that provides minimal benefit at a very high cost.

The general problem is larger than that addressed in this section. Disputes over whether a patient's medical insurance covers an expensive treatment protocol that is nonetheless clearly beneficial for the patient are addressed in Chapter 1.D.2, and the general

problem of designing medical insurance that allocates resources rationally and ethically is treated in Chapter 1.D.1. We here address a more limited but related issue: may a health care provider resist treatment on the grounds that it would confer insufficient medical benefit.

On one level the issue seems easy. Every diagnosis necessarily implies a decision to consider some treatments and rule out others. There is no reason to remove the appendix when the diagnosis is stomach flu, and it would be nonsense to argue even that the physician has an ethical duty to discuss the surgical option with the patient, much less to comply with the patient's request for an appendectomy.

But other cases are less clear. We might move one step on the implausible-plausible continuum by examining the facts of In re Baby K., 16 F.3d 590 (4th Cir. 1994), in which the mother sought artificial ventilation for her anencephalic newborn. The baby could breathe on her own for limited periods, and was transferred to a nursing home, but was brought to the hospital emergency room whenever respiratory crises developed. The doctors believed that providing respiratory assistance was medically and ethically inappropriate for a baby with no cerebrum who was on that account likely to die soon in any event and who, while alive, was permanently unconscious and

without the ability to see, hear, think, develop any self-awareness or awareness of her environment. The court nonetheless held the hospital obliged to provide the treatment under EMTALA (the Emergency Medical Treatment and Active Labor Act), an almost certainly unintended result of the legislation that nonetheless

follows logically from a mechanical application of its terms. The required breathing assistance was well within the hospital's capabilities and would generally be provided to any emergency department patient with breathing difficulties, as part of the ordinary and required "stabilization" of the patient's condition. These conclusions, necessary to the court's application of EMTALA, also tell us that while many would agree with the physicians' judgment in this case, that judgment is not based upon the treatment's inability to extend life, and thus the treatment is not literally "futile" in the way that an appendectomy is a futile treatment for stomach flu. Invoking futility in this case is instead based upon the view that treatment is inappropriate when it will achieve only a limited extension of life and there is no chance that the life will have any meaning for the patient.

One can thus see that the debate over medical futility is often a debate over values rather than over medical facts. Consider, for example, In re Wanglie, 2 BioLaw U:2161 (Aug.-Sept. 1991) (Minn. 4th Dist. Ct.) in which the family sought continued respiratory support for an 86 year old diagnosed as irreversibly unconscious. Does it matter whether there is some non-zero chance, however slight, that the patient might recover consciousness if kept alive? Does it

matter if it is clear that the patient herself wanted to be kept alive, even if permanently unconscious, because she valued even that limited an existence?

The same issues arise in connection with Do-Not-Resuscitate (DNR) orders, although the question is often put differently in form because consent to resuscitation is ordinarily presumed. In the case

of DNR orders, then, consent is needed to *forego* resuscitation, rather than to provide it, and usual protocols contemplate a discussion of the question between the physician writing the DNR order and the patient or other person who is the patient's medical decision-maker. The question arises when consent to a DNR order is refused despite the physician's belief that resuscitation would be inappropriate. Consider, for example, In re Jane Doe, 418 S.E.2d 3 (Ga. 1992), in which the father refused to consent to a DNR order for his 13-year-old daughter, who was suffering from a degenerative nerve disease and "vacillated between stupor and coma." The hospital believed that aggressive medical treatment would constitute abuse, but the court held that under state law, CPR could not be withheld unless both parents agreed to a DNR order. Resuscitation would thus take place if the girl's heart stopped beating. The court did not consider the futility issue, rendering instead a mechanical application of the state statute. Yet note that if the question were whether to perform surgery for an unrelated condition—imagine, for example, that the girl also developed kidney failure, which might be treated with a transplant from her brother—the physician might simply decide against it without discussion with the parents, and certainly

without seeking their permission for the surgery. The statute relied upon in *Doe* would have been inapplicable. Would the physician's decision be ethical? If so, then how does one distinguish the resuscitation required in *Doe*?

Defining Futility. Futility is generally analyzed under two rubrics: qualitative and quantitative futility. Under qualitative futility, the claim is that medical treatment cannot provide a sufficient benefit to justify its use. Some commentators argue that

there is qualitative futility when the patient would not recover enough to go home from the hospital, some find qualitative futility when the patient is permanently unconscious, and others believe there is futility only when treatment cannot provide a physiological benefit. In this third view, treatment is not futile as long as it can prolong life or affect the quality of life in any way.

Under quantitative futility, the claim is that there is too low a likelihood that medical treatment will have its desired effect. Here, too, there is disagreement as to when futility exists. Some would find futility when a particular treatment has been consistently unsuccessful for at least 100 tries. Lawrence J. Schneiderman et al., *Medical Futility: Its Meaning and Ethical Implications*, 112 Ann. Intern. Med. 949 (1990). Others would place it at different likelihoods of success. In one study, researchers discussed with internal medicine residents the reasons why the residents wrote do-not-resuscitate (DNR) orders for their patients. In two-thirds of the cases in which quantitative futility was

a contributing factor to the DNR order, the likelihood that the patient would be resuscitated and be able to go home from the hospital was 1 percent or less. In 9 percent of cases, on the other hand, the likelihood of success was 20 percent or more. J. Randall Curtis et al., *Use of the Medical Futility Rationale in Do-Not-Attempt-Resuscitation Orders*, 273 JAMA 124, 126–127 (1995). Of course, one's view about quantitative futility depends on the qualitative benefit to be gained. The greater the potential qualitative benefit, the lower the likelihood of benefit before a treatment would be considered futile.

Ultimately, it appears that the futility debate may be misnamed. There is little controversy over the small group of cases in which there is true medical futility in the narrow physiological sense, as in our appendectomy example. And the larger group of more difficult cases are not about true medical futility but instead involve value judgments that must share basic principles with the system we adopt generally for allocating health care resources.

Futility Cases. There has been only a smattering of reported legal decisions concerning the denial of requested care on grounds of futility, and courts so far have split between siding with the patient or the patient's family and siding with the physicians and hospital, so it is difficult to draw firm conclusions about the state of the law.

Nevertheless, at least one pattern seems to have emerged: courts are more likely to approve decisions by physicians to deny care than to decide themselves to deny it. In the *Causey* case, for example, the court

essentially held that futility decisions already made and implemented are permissible as long as they are consistent with the professional standard of care. Causey v. St. Francis Medical Center, 719 So.2d 1072 (La. Ct. App. 1998). On the other hand, *Baby K* illustrates that courts are less inclined to grant a hospital or physician prospective authorization to withhold life-sustaining treatment on grounds of futility. The difference, perhaps, is between declining to second-guess physicians who taken the responsibility of invoking futility in difficult circumstances, and taking the responsibility for making the decision in the first place.

Note that in cases upholding a futility decision, the physicians always sought input from the hospital ethics committee and gave

the family time to arrange for transfer of the patient to another facility. In other words, the physicians and hospitals followed good procedural standards before withdrawing treatment.

Futility Statutes. Several states have addressed futility by statute. See, e.g., Ark. Code § 20–6–109; Tenn. Code § 68–11–1808; Tex. Health & Safety Code Ann. § 166.046; Va. Code Ann. § 54.1–2990.

The Texas statute is notable for the procedures it sets out to invoke futility. In Texas, a physician's refusal to honor a request for treatment requires review by an ethics or medical committee, and the patient or surrogate decision-maker is entitled to attend the committee's meeting and receive a written explanation for the committee's decision. If the patient or family disagrees with a committee's decision affirming the refusal, the physician must

make a reasonable effort to transfer the patient to a physician who will provide the desired treatment. Although treatment must be provided pending transfer, there is no obligation to provide treatment beyond the tenth day after a committee decision in favor of refusing treatment.

While some view the Texas approach as a model, it has provoked considerable controversy in some cases when doctors and hospitals have invoked it, and family members have been able to delay the withdrawal of treatment through court challenges.

While Texas clearly recognizes futility as a basis for withholding treatment, other statutes tend to authorize withholding treatment only under a very narrow or uncertain definition of futility or require the provision of treatment pending transfer to a facility that is willing to provide care.

Virginia's statute, Va. Code Ann. § 54.1–2990, seems to clearly authorize physicians to withhold care they deem futile, and some hospitals have implemented futility policies without incident. Other hospitals in the state, however, see uncertainty in the statutory authorization, and there have been efforts to clarify the law. Similarly, in other states, such as Arkansas and Tennessee, whose statutes seem to authorize physicians to withhold care they deem futile, some hospitals are comfortable implementing the statutory authorization while other hospitals are not.

Futility and Brain Death. Although futility has been debated seriously as an issue only since the late

1980s, the development of the concept of brain death in the late 1960s and early 1970s can be seen as an early manifestation of the futility concern.

As we saw in Chapter 6, death was traditionally pronounced based on the cessation of the beating of the heart and the breathing of the lungs. In other words, death was determined on the basis of "cardiopulmonary" criteria. With advances in medical technology, however, it became possible to support a person's heart and lungs with mechanical ventilation even after cessation of brain function. According to cardiopulmonary criteria, these persons were not dead, but to many observers, they seemed to have lost their vitality. Patients who have lost all brain function appear to be in a very deep coma. Because of the profound loss of functioning, the permanence of the condition, and the inability to prolong the state for very long, many observers questioned whether it made sense to provide treatment to these patients.

Ordinarily, such questioning might lead to a right to have medical treatment withdrawn when a person's brain stops functioning, just as Karen Quinlan's persistent vegetative state led the New Jersey Supreme Court to recognize her right to have her ventilator withdrawn. Instead, "brain" criteria for death were developed. In a report that was highly influential in leading to the acceptance of "brain death," the authors wrote:

> Our primary purpose is to define irreversible coma as a new criterion for death. There are two reasons why there is need for a definition: (1)

> Improvements in resuscitative and supportive measures have led to increased efforts to save those who are desperately injured. Sometimes these efforts have only partial success so that the result is an individual whose heart continues to beat but whose brain is irreversibly damaged. The burden is great on patients who suffer permanent loss of intellect, on their families, on the hospitals, and on those in need of hospital beds already occupied by these comatose patients. (2) Obsolete criteria for the definition of death can lead to controversy in obtaining organs for transplantation.

Report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death, A Definition of Irreversible Coma, 205 JAMA 85, 85 (1968).

As this excerpt suggests, the adoption of brain criteria for death may have been motivated primarily by futility-type concerns—the high burdens and low benefits of treatment—as well as by concerns about the shortage of organs for transplantation. Philosophical

considerations about the meaning of life seemingly played little role in the change.

[1] If a prisoner is refusing treatment for the same kinds of reasons as non-incarcerated patients, then courts generally uphold the refusal.

[2] This factor is relevant because of the medical evidence in the record that people of Saikewicz's age do not tolerate the chemotherapy as well as younger people and that the chance of remission is decreased. Age is irrelevant, of course, to the question of the value or quality of life.

[3] Miller, Dresser & Kim, Advance euthanasia directives, 45(2) Journal of Medical Ethics 84 (2019), https://jme.bmj.com/content/45/2/84.

CHAPTER 8

SELECTED ISSUES IN REPRODUCTIVE MEDICINE

A. THE RIGHT TO PROCREATE

Though it isn't mentioned in the Constitution, the right to procreate is firmly established as a fundamental right. In Skinner v. Oklahoma, 316 U.S. 535 (1942), the Supreme Court wrote that

> We are dealing here with legislation which involves one of the basic civil rights of man. Marriage and procreation are fundamental to the very existence and survival of the race. Id. at 541.

The Court has never identified the full extent of the right to procreate. In *Skinner*, at issue was an Oklahoma law which called for the sterilization of some persons after their third conviction for a felony. The Court rejected a power of the government to sterilize people without their consent—the state cannot permanently deprive individuals of their ability to reproduce. But what if the government interferes in other ways with procreation? Can the state prohibit commercial surrogacy, thereby limiting the ability of infertile couples to have children?

The Supreme Court has objected to policies that impose a "heavy burden" on the right to procreate in the absence of a substantial justification for the policies. For example, in Cleveland Board of

Education v. LaFleur, 414 U.S. 632 (1974), the Court invalidated a maternity leave policy that required teachers to go on leave at least four months prior to their due date for the birth of their child.

At the same time, the Court permits burdens on the right to procreate that are "marginal and indirect." Thus, in Dandridge v. Williams, 397 U.S. 471 (1970), the State of Maryland could cap welfare benefits for poor families with children once the family size reached six people. In upholding the cap on benefits, the Court's decision presaged its abortion funding decisions, in which it has held that while the government cannot prevent a woman from seeking an abortion, it does not have to fund her abortion.

The case law so far indicates that the government has considerable freedom to withhold assistance from people trying to procreate but less freedom to interfere when people are trying to procreate without state support. Still, it is likely that some kinds of interference are permissible, as, for example, when states prohibit commercial surrogacy. Or states likely could limit the number of embryos transferred to a woman's uterus during in vitro fertilization (IVF) because of the risks to maternal and child health from multiple-infant births (e.g., triplets). But if a state were to ban IVF, the Court likely would view that as unconstitutional. (In vitro fertilization means fertilization "in glass" rather than in the woman's body, which would be "in vivo" fertilization.)

In assessing limits on the right to reproduce, the *Skinner* Court's reasons for striking down the Oklahoma sterilization law should be relevant. As

mentioned, the Court wrote that procreation is fundamental to societal existence and survival. In addition, the Court worried that a

societal majority could use a power to sterilize to discriminate against disfavored groups. Finally, the Court observed that sterilization was irreversible.[1] 316 U.S. at 541. But while relevant, these considerations are not decisive. For example, Maryland's cap on welfare benefits likely reflected the desire of a majority of the public to discourage procreation by poor families.

B. ASSISTED REPRODUCTION

As we will see, the law sometimes tries to limit access to assisted reproduction. More commonly, the legal debates are about parentage—when donors of eggs or sperm do not intend to raise the children that result, what must they do to avoid legal responsibilities for the children? And what happens if they change their minds and want to assume the role of the child's parent, as in the case of a surrogate mother or donor of eggs or sperm? With evolution in the use of assisted reproduction, states have revisited traditional legal rules attributing parentage, sometimes through legislative action, other times via judicial review. With regard to the latter, the constitutional right to parent one's offspring has been invoked by some courts to protect the right of sperm donors to establish their fatherhood.

1. ARTIFICIAL INSEMINATION BY DONOR (AID)

a. Basic Rules

Artificial insemination is an older medical procedure whose availability long predates the modern cases that have drawn most of the public attention to assisted conception issues. It traditionally was employed by married couples in which the husband was infertile but reproduction was possible with sperm from another

man. With the development of intracytoplasmic sperm injection (ICSI), artificial insemination can be performed with a single sperm cell, significantly reducing the infertility rate among men. Still, some men may not be able to father a child even with ICSI or its costs may make it unaffordable, and the couple may turn to sperm donation. When a married woman is impregnated artificially with a donor's sperm, state law typically provides that the woman's husband, not the sperm donor, is the child's legal father, either by case law (People v. Sorensen, 437 P.2d 495 (Cal. 1968)) or by statute (see § 5 of the 1973 Uniform Parentage Act, § 705 of the 2002 and 2017 Acts). (We ignore cases in which the husband's sperm is used, as they raise no issues of parentage attribution.)

As artificial insemination began to be used more frequently by single women or women with same-sex partners, things became more complicated. The 1973 Uniform Parentage Act (UPA) does not address the case of an artificially-impregnated woman who is not married and who therefore has no husband to whom

the law may assign paternal responsibilities. In states following the 1973 UPA, the sperm donor may become the legal father, or the child will have no legal father. The 2002 and 2017 UPAs extended the Act's application to unmarried women, and the revised Acts state that donors are not parents of children conceived through assisted reproduction (§ 702). In addition, the recognition of same-sex marriage has addressed much of the parentage question by eliminating the disparate treatment of opposite-sex and same-sex relationships.

When unmarried women use artificial insemination, a few considerations matter. Under §§ 702 and 703 of the Uniform

Parentage Acts (UPAs) of 2002 and 2017, a sperm donor does not become a legal parent unless he donates "with the intent to be" the father of the child. Under § 5 of the 1973 UPA, the donor did not become a parent if the insemination was supervised by a licensed physician. State statutes generally reflect the principles in the UPAs, so the cases tend to depend primarily on the intent of the donor and the mother, with some consideration of whether the insemination occurred under the supervision of a physician. For example, one approach seen among states is to presume that the donor is not the father when a physician performs the insemination, but that the donor is the father when a physician is not involved. In either case, the donor and the mother can agree to override the presumption. (Note that in some states, when the donor is not the spouse of the woman, artificial insemination must be performed by a physician or

someone under the supervision of a physician. See, e.g., Ohio Stat. § 3111.90).

A number of cases illustrate the general approach. For an example of the role of intent, consider the application of a Kansas law that adopts the presumption that if a physician performs the insemination, the donor does not become the child's father unless the donor and mother agreed to his paternity in writing. Kan. Stat. § 23–2208(f). In In re K.M.H., 169 P.3d 1025 (Kan. 2007), the donor provided sperm to a friend for insemination by a physician. After the birth of twins, the donor asserted paternity. Because there was no written agreement to his paternity, the court denied his claim.

As in other areas of the law, statutory interpretation can be complicated. The donor in *K.M.H.* tried to invoke the state's general provisions for establishing paternity, which included as a basis if a "man notoriously or in writing recognizes paternity of the child." The court rejected that argument on the basis of the traditional principle of statutory interpretation that a "specific statute controls over a general statute." But a California court recognized paternity when it allowed a sperm donor to invoke the state's general paternity statute.

In Jason P. v. Danielle S., 171 Cal.Rptr.3d 789 (App. 2014), the facts were similar to *K.M.H.* The donor provided sperm to a friend for insemination by a physician, and there was no written agreement that

he would be the child's father.[2] California has the same paternity provision as Kansas. If artificial insemination is performed by a physician, the donor does not become the child's father unless the donor and mother agreed to his paternity in writing. Cal. Fam. Code § 7613(b)(1). As in *K.M.H.*, the requirement of a written agreement could have barred the donor's paternity.

But the *Jason P.* court decided that the general paternity statute also should be considered. More specifically, the court invoked a provision of the state's UPA that recognizes paternity when a "presumed parent receives the child into his or her home and openly holds out the child as his or her natural child." Cal. Fam. Code § 7611(d). For two-and-a-half years after the birth of the child, the donor had been involved in the child's life, and the child called him "Dada." But then, the mother ended her relationship with the donor. In turning to the general paternity provisions, the

California court effectively held that when deciding whether the mother and the donor intended him to be the father of the child, intent should be judged by their conduct after the child's birth as well as beforehand. See also Thomas S. v. Robin Y., 618 N.Y.S.2d 356 (App.Div. 1994) (donor recognized as father under general paternity law principles when he had a social relationship with the child).

As cases have arisen, and gaps in the law have been exposed, state legislatures often have adapted. Consider, for example, the traditional rule that sperm donors do not become fathers when a physician performs the insemination. Some states have maintained that rule. Thus, in Bruce v. Boardwine, 770 S.E.2d 774 (Va. App. 2015), a donor was able to establish paternity after donating to a friend because the mother performed the insemination herself. In California, on the other hand, the legislature modified its traditional rule following court decisions. In a 1986 case, Jhordan C. v. Mary K., 224 Cal.Rptr. 530 (App. 1986), a man asserted parental status after donating sperm to an unmarried lesbian couple. The mother testified that she and the donor did not intend for him to have ongoing involvement with the child. The sperm donor testified that their intent was for him to play a paternal role. Because the artificial insemination occurred in the mother's home without the participation of a physician, the statutory provision denying fatherhood did not apply, and the court recognized the sperm donor as the father. Under current California law, if sperm donation occurs without the participation of a physician, the donor is still presumed to be the father, but not if the donor and mother agree prior to conception that the donor will not be the father. Cal. Fam. Code § 7613(b)(2).

The California legislature also filled the gap for cases in which sperm donation occurs with the participation of a physician. In Steven S. v. Deborah D., 25 Cal.Rptr.3d 482 (App. 2005), the sperm donor was unable to assert paternity because of the same

California law that established paternity for the donor in *Jhordan C.* In *Steven S.*, the donor lost because he and the mother had used a physician to perform the insemination. Now in California, as indicated above, the donor and mother can agree in advance that he will be treated as the father (with the *Jason P.* court also allowing for post-birth conduct to establish paternity).

Some courts have required an opportunity for the sperm donor and woman to establish intent even when state statutes did not provide for that. In McIntyre v. Crouch, 780 P.2d 239 (Or. App. 1989), a man donated sperm to an acquaintance who inseminated herself without physician involvement. Under Oregon law at the time, sperm donors were not recognized as fathers. Nevertheless, the court cited U.S. Supreme Court cases involving the constitutional rights of unmarried fathers to hold that the donor should be recognized as the child's father if, after remand, he could establish that he and the mother "agreed that he should have the rights and responsibilities of fatherhood and in reliance thereon he donated his semen." See also C.O. v. W.S., 639 N.E.2d 523 (Ohio C.P. Cuyahoga 1994).

Ferguson v. McKiernan, 940 A.2d 1236 (Pa. 2007), provides another important example of judicial reliance on the intent of the donor and recipient. In that case, the donor provided sperm to a friend and former paramour, who used the sperm to have children. The woman had previously undergone a tubal ligation, so the

insemination required in vitro fertilization (IVF) and therefore physician

388

involvement. Although the donor and the mother had orally agreed in advance that he would not be treated as the father, she filed for child support when her twins were five. Pennsylvania had not adopted legislation governing artificial insemination and paternity, and the mother invoked a general paternity statute according to which children born out of wedlock shall be treated as if their biological parents were married. The court rejected that argument, holding instead that since the sperm donation functionally was like that of an anonymous donation to a clinic, the donor should not be treated as the father of the children.

Even when a court seems to override donor and mother intent, intent may still prevail. In E.E. v. O.M.G.R., 20 A.3d 1171 (N.J. Super. Ct. Ch. Div. 2011), the donor provided sperm to a single female friend, and she inseminated herself at home without a physician's participation. The donor and mother had signed a written agreement that he would voluntarily relinquish all parental rights and responsibilities. When they sought a court order terminating his parental role, the court declined the request. Under the state statute, sperm donors are not considered fathers (absent an agreement to the contrary) when a physician performs the insemination. According to the court, the failure to use a physician prevented the mother and donor from invoking the statute. As a practical matter, though, intent mattered. The court awarded sole custody to the mother with no parenting time for the donor, accepted the mother's decision not to seek child

389

support, and left open the possibility of terminating the donor's parental rights at a later date.

Of course, it's simpler when intent and the statutory language align. In Lamaritata v. Lucas, 823 So.2d 316 (Fla. App. 2002), the donor and mother had signed a contract under which he would have no parental rights or responsibilities. In addition, under the relevant Florida statute, sperm donors do not become fathers absent an agreement to the contrary. Accordingly, the donor's assertion of paternity failed.

b. Use by Lesbian Couples

In many of the AID cases involving an unmarried woman, the mother was part of a lesbian couple who intended to raise the resulting child together, and who did not wish the sperm donor to have a parental relationship with the child. They sometimes had an explicit agreement to that effect with the donor. Prevailing law, however, often denied enforcement to such pre-birth disavowals of parenthood, and sometimes the lesbian couple themselves undercut the agreement by permitting the donor contact with the child. Whether because of this contact, *Thomas S.*, *supra,* or as a result of the rules already surveyed, *Jhordan C., supra*, donors in such cases were sometimes recognized as the child's father despite the parties' agreement. Of course, as the discussion of cases indicates, problems with donor assertions of fatherhood could be avoided by using a fertility clinic and an anonymous donor.

But societal biases, including same-sex marriage bans, often made that difficult. Medical personnel

traditionally declined to perform inseminations of unmarried women, and a 1988 study found that 61 percent of physicians said they would reject requests for AID from an unmarried woman "without a partner." Attitudes have evolved over time. In a 2005 survey of directors of assisted reproduction programs, 20 percent reported that they would reject requests for AID from a woman without a husband or partner, and in 2006, the American Society for Reproductive Medicine wrote in an ethics committee report that fertility clinics "should treat all requests for assisted reproduction equally without regard to marital status or sexual orientation."

With recognition of a right to marry for same-sex couples, many questions about parentage disappear. For example, state statutes typically recognize the husband of the mother as the father of a child born after artificial insemination, if the husband consented to the insemination. The U.S. Supreme Court has held that these statutes also must recognize the same-sex spouse of the mother as a parent of the child, foreclosing the need for the mother's spouse to adopt the child. Pavan v. Smith, 137 S. Ct. 2075 (2017).

c. Newer Issues

Just as states have made it easier for adopted children to identify their biological parents, there is increasing interest in making it easier for children to find out the names of the anonymous sperm (or egg) donors who participated in their procreation. In 2011, the State of Washington passed a law under which

children are entitled to the names of gamete donors, unless the donor chooses to remain anonymous, Rev. Code Wash. § 26.26A.820. Section 904 of the 2017 UPA incorporates the Washington approach, and it also has been adopted in California. In

addition to making identification available unless the donor opts for anonymity, clinics are required to make a good faith effort to notify donors of their option to revoke their choice of anonymity when their children seek identifying information.

An important question regarding the trend toward identification is whether it will discourage men from becoming sperm donors.

2. SURROGATE MOTHERHOOD, EGG DONATION, AND IN-VITRO FERTILIZATION

In a traditional, simple egg donation, an otherwise normal woman who is incapable of producing fertile eggs can nonetheless become pregnant with her husband's child, by using his sperm to fertilize a donated egg before implantation in her uterus. This process is analogous to classic AID, which allows an infertile man to be the legal father of a child born to his wife and is a common use of IVF. It also is more straightforward in terms of parentage since egg donation requires the services of a physician.

Surrogate motherhood refers to procreation in which the "surrogate mother" provides her gestational services to another woman or couple, who intend to raise the child. In the original and simplest form of surrogate motherhood, the child is the genetic child of the surrogate, who has typically become

impregnated through artificial insemination with the sperm of the intended mother's husband (traditional surrogacy). Alternatively, an egg from the intended mother is fertilized in vitro, typically with the sperm of her husband, and then implanted in the surrogate's uterus (gestational surrogacy).

Other variations are possible. For example, the egg may come from a woman other than the intended mother and the surrogate mother. The technologies of egg retrieval, IVF, and implantation permit complete separation of the identities of sperm donor, egg donor, gestational mother, partner of the gestational mother, and intended mother and father, so that up to six persons, all possibly unmarried, can have some relationship to the resulting child.

a. Surrogacy Contracts Under Traditional Law

While traditional law clearly did not contemplate such surrogate arrangements, many existing principles seem to bear on them. Most of the potentially applicable law comes from the field of adoption, and poses barriers to the enforcement of any contract between the surrogate and the intended adoptive mother.

(1) Regulation of Private Placement Adoptions

Most American states allow "private placement" adoptions in which the biological parents make arrangements with the adoptive parents themselves, either directly or through an intermediary such as a lawyer. A judicial proceeding is still required to formalize the adoption and establish the adoptive

parents' rights, and in that proceeding, the judge will receive evidence to establish that the adoptive parents meet required standards of fitness. In addition, most states regulate any payments made by the adoptive parents, either to the biological mother who relinquishes her child for adoption, or to the lawyer or other intermediary. The details of these laws vary, as to both substance and procedure, but most share a common theme. Intermediaries may not collect fees for securing the child or the mother's consent;

this means, for example, that an attorney intermediary may collect fees for actual legal work done, but not for services as a child broker. *E.g.,* Galison v. District of Columbia, 402 A.2d 1263 (D.C. App. 1979). Biological mothers may be compensated for expenses they incur as a result of the pregnancy, such as medical bills and, depending upon the details of state regulation, possibly lost earnings or basic living expenses, but they may not be paid a fee in exchange for their consent to the adoption. Such a fee might be considered unlawful baby-selling.

It is clear that such rules pose potential problems for the traditional surrogacy contract, in which the surrogate bears a child conceived through artificial insemination using the sperm of the intended mother's husband. The fee paid to the surrogate is, in part, payment for the consent to adopt, even though it is also compensation for undergoing the pregnancy and delivery. That is clear because the intended mother enters the agreement in order to receive the child for adoption, and often final payments are not made until custody is transferred to her and her husband. The willingness of the surrogate to permit

her adoption of the child is therefore an essential part of the agreement. The fee promised by the agreement may thus be illegal. In some states all fee arrangements made in connection with an adoption must be disclosed in the adoption proceedings and approved by the court; clearly fees paid in exchange for the consent would not be approved.

The requirement that a court review and approve the fitness of the adoptive parents also can interfere with surrogacy arrangements. Court review means that completion of the adoption is not entirely

in the parties' control. To be sure, as implemented in most private placement adoptions, this process would not usually cause a problem with surrogacy so long as the parties remain in agreement, because the court's evaluation of the adoptive parents is not ordinarily comparative: The court does not ask whether better parents could be found for the child, but only whether these parents are acceptable, and most intended adoptive mothers under surrogacy agreements surely meet this test. Nonetheless, there may be surrogate contract arrangements that would transfer a child to a mother who is clearly unsuitable. Moreover, if the parties do not remain in agreement, the surrogate, as the legal mother at birth, will have priority over the intended mother. Most states do not recognize consent to adoption given by the mother prior to the birth of the child. The surrogate mother is thus free to repudiate any consent to adoption she may have given in the pre-birth surrogate contract.

395

(2) Rights and Identity of the Father

While the surrogate may be able to repudiate the consent to adoption of her child that she gave in a surrogacy contract, she cannot thereby exclude claims by the father of her child. In such a case, if the father promptly asserts his rights, he will transform the dispute into an ordinary custody contest. Given the relative situations of the typical surrogate and the typical father, one might expect the father to win such contests most of the time. The father usually has a stable marriage; the mother may not be married. The father is typically financially secure, as the cost of surrogate agreements for the adoptive parents will usually exclude those who are not; the mother is often financially insecure, which may be why she agreed to serve as a surrogate mother in the first place. The

differences in economic status will usually be correlated with differences in educational attainments and employment status, with the result that the father's overall situation will be more favorable for the child than the mother's.

One possible difficulty for the father may arise from the rules governing AID, reviewed above, under which the surrogate's husband, if she has one, rather than the semen donor, might be considered the legal father of the surrogate's child. To forestall such results, surrogacy agencies typically require the written agreement of the surrogate's husband, where she has one, but the legal effect of such agreements has not been tested. Of course, AID statutes were not written with surrogacy agreements in mind, and a

court might therefore choose not to apply the statute to such a case.

(3) In re Baby M.

The difficulties of enforcing surrogacy contracts under traditional law are well illustrated by an early but still well-known surrogate mothering case, In re Baby M., 537 A.2d 1227 (N.J. 1988). Elizabeth Stern, a physician, had multiple sclerosis, and was unwilling to go through a pregnancy because of the risk to her health which she believed would result. Nonetheless, she and her husband, William Stern, wanted children. William contracted with Mary Beth Whitehead to bear his child, the two having been brought together by the Infertility Center of New York, which specialized in finding surrogate mothers for infertile couples. Mrs. Whitehead was married and had children. She had previously offered to be a surrogate for another couple, but never became pregnant. Artificial insemination with Mr. Stern's sperm was successful, however. After the child was born, Mrs. Whitehead had

great doubts about her ability to surrender it; among other things, she felt the baby girl closely resembled her other daughter. Despite her doubts, she surrendered the child to the Sterns three days after the birth. But that evening she called the Sterns in despair. Afraid that Mrs. Whitehead might commit suicide if she did not see the child again, the Sterns agreed to return her to Mrs. Whitehead for a week.

In fact, the Sterns did not recover the baby until four months later, when she was forcibly removed

from her grandparents' home in Florida, to which Mrs. Whitehead had fled with the child. From that time forward, Melissa, as the Sterns called her, lived with the Sterns, while a court battle over her ensued. The Sterns sought enforcement of the surrogacy contract; they also alleged that the child's best interests supported placing her with them.

The contract had three parties: William Stern, Mary Whitehead, and Mary's husband, Richard. Mrs. Whitehead had agreed to become pregnant, carry the child, bear it, deliver it to the Sterns, and do whatever was necessary to terminate her parental rights so that Elizabeth Stern could adopt the child. Mr. Whitehead promised to do whatever was necessary to rebut the presumption of his paternity that would arise under New Jersey law, since the child would be born to his wife. Mr. Stern agreed to pay Mrs. Whitehead $10,000 upon delivery of the child to him, and also to pay the Infertility Center $7500 for its services, which were to include completing the adoption process. Mrs. Stern was not a party to the agreement, but it provided that in the event of Mr. Stern's death, she would have sole custody.

The court concluded that while all parties had entered the agreement in good faith, it was invalid under New Jersey law. The specified payments to both Mrs. Whitehead and the Infertility Center were found to violate New Jersey statutes regulating the fees that could be charged in connection with adoption. The court rejected the claim that the fees to Mrs. Whitehead were only for "services,"

398

concluding instead that they were prohibited payments for her consent to Mrs. Stern's adoption of the child, and it rejected the claim that the fees paid to the Infertility Center were only for legal services. In reaching these conclusions the court looked beyond the characterization of these payments contained in the agreement, focusing instead on what it viewed as the reality of the exchange.

The court also found that an agreement to surrender a child not yet born was not enforceable under New Jersey law because it did not comply with any of the permissible New Jersey procedures under which parental rights can be terminated. "[A] contractual agreement to abandon one's parental rights, or not to contest a termination action, will not be enforced in our courts." *Id*. at 1243. If Mrs. Whitehead's parental rights were not terminated, then the adoption could not go forward. It also found that there was nothing in the record that would justify an involuntary termination of Mrs. Whitehead's parental rights, given the well-established family law principle that a parent's rights cannot be terminated merely because a court judges that a better parent is available. Parents whose fitness meets minimum standards have a protected interest in retaining their parental rights as against a stranger, even one who seems superior.

With the contract effectively voided, the case turned into a custody contest between the girl's legally recognized parents, Mr. Stern and Mrs. Whitehead. The court held that the contest should be judged by the traditional custody decision rule:

placement is determined by the child's best interests, giving no weight to the surrogacy contract. In comparing the Whitehead and Stern homes, the court found, not surprisingly, that the evidence strongly favored the Stern's. While granting Mr. Stern primary custody, the court also held that Mrs. Whitehead was entitled to visitation rights, a result which followed almost necessarily from the court's decision that her parental rights continue. The trial court later ruled that Mrs. Whitehead would be entitled to 8 hours of unsupervised visitation per week, to increase over the course of the year to two days every two weeks, including overnight.

Baby M. is a straightforward application of traditional rules. While its analysis was followed by most courts, at least one held that the anti-baby selling provisions contained in adoption and custody laws were not intended to apply to surrogacy arrangements, and thus did not bar surrogacy contracts. Surrogate Parenting Associates v. Kentucky, 704 S.W.2d 209 (Ky. 1986). Although the court did not find surrogacy contracts void, it did hold that the contracts were unenforceable, so the surrogate could choose whether to relinquish or retain her parental rights. After the case, the Kentucky legislature passed a statute making surrogacy contracts illegal.

Note that most surrogate parenting contracts are carried out successfully, in that the parties do not later disagree, and for them

the refusal of courts to enforce the agreement may not matter. Even then, however, to complete their intended transaction, the

surrogate will have to execute, post-birth, a document relinquishing her parental rights, and judicial approval of the intended mother's adoption will be required. But even if the surrogacy contract is not valid in the sense that it would not be enforced over the surrogate's objection, where there is no objection, courts are likely to approve the adoption despite the fact that the child's existence results from the invalid contract. E.g., Adoption of Baby A and Baby B, 877 P.2d 107 (Or. App. 1994).

There was considerable legislative activity after *Baby M.*, and today nearly half the states have statutes regulating surrogacy agreements. Only a few criminalize the making of such agreements; most simply declare the contract void, a result consistent with traditional law. A handful recognize surrogacy contracts, but most of these combine that recognition with important limitations, such as disallowing payments to the surrogate that exceed her expenses, or allowing the surrogate to revoke her agreement and keep the child if she acts promptly after birth. For a more complete description of the various state laws, see Columbia Law School Sexuality & Gender Law Clinic, *Surrogacy Law and Policy in the U.S.* (May 2016).

One might wonder about the constitutionality of surrogacy bans. Do they violate the right to procreate? In the *Baby M.* case, the court considered whether Mrs. Stern was treated differently than a man who uses assisted reproduction to have a child with his wife. As we have seen, if a couple uses sperm from a donor to have a child, the husband of the

mother is recognized as the father of the child. Similarly, argued the Sterns in *Baby M.*, if a couple uses an egg from a donor to have a child, the wife of the father should be recognized as the mother of the child.

The *Baby M.* court acknowledged that this argument had force in the case of a donor who provided an egg and nothing else. It would make sense to have the same rules for egg donors and sperm donors. But Mrs. Whitehead did much more than provide an egg—she carried the pregnancy for nine months, and that gave her a much greater claim of parenthood than a man who donates sperm. Hence, wrote the court, there was a valid basis for treating traditional surrogacy differently than sperm donation. 537 A.2d at 1254–1255.

b. Gestational Surrogacy and Changing Law

Whether or not one believes it invokes the appropriate policy, the traditional law applied in *Baby M.* yields a clear result where the surrogate is the biological mother of the child in question. More confused, however, is the case in which the surrogate is the child's gestational mother but the intended mother is the child's genetic mother. This occurs when a fertilized egg of the intended mother is implanted in the uterus of the surrogate, who then carries the child to term. In the event of a dispute, which woman should the law recognize as the child's legal mother? Under these facts, one is tempted to conclude that the genetic mother should prevail. Yet consider the equally possible case in which the

gestational mother is the intended mother—where, unable to produce her own fertile eggs, she carries to term a donated egg that was fertilized in vitro with her husband's sperm. Should the egg donor then be able to claim the child is hers?

Faced with such possibilities, the California Supreme Court concluded that both gestational and genetic mothers satisfy the traditional legal standards for "mother" contained in statutes like the Uniform Parentage Act, since motherhood under those pre-technology standards could be shown by either genetic proof or by evidence of having given birth to the child. Declining to give both women the status of legal mother, the court concluded that the question should be decided according to the intent of the parties at the time of the child's conception. It therefore held that the surrogate, who in the case before it had carried to term the intended mother's genetic child, was not the child's legal mother, and must therefore surrender it to the intended mother and her husband (whose sperm had been used to fertilize his wife's egg in vitro). Johnson v. Calvert, 851 P.2d 776 (Cal. 1993).

If the parties' intent had been the test applied in *Baby M.*, then the surrogacy contract there would have been effectively enforceable. But California has not applied the *Johnson* intent test to traditional surrogacies of the *Baby M.* kind. See Moschetta v. Moschetta, 30 Cal.Rptr.2d 893 (App. 1994) (holding that the intent of the parties was relevant only when the gestational mother is not the genetic mother, and that a surrogacy agreement could not therefore be

enforced over the opposition of the surrogate who was both the gestational and genetic mother). On the other hand, the intent test was employed in a relatively bizarre case at the other extreme. In

Buzzanca v. Buzzanca, 72 Cal.Rptr.2d 280 (App. 1998) the parties, while married, had a fertilized egg implanted in the uterus of a surrogate who had agreed to carry the child to term. Neither husband nor wife was genetically related to the child. When husband and wife filed for divorce before the child's birth, the court had to decide whether this was a child "of the marriage." The husband denied all rights and responsibility to the child, as did its gestational mother. The wife asserted she was the child's legal mother, and claimed primary custody as well as child support from her husband as its legal father. The genetic parents were not parties to the case and were apparently unknown. The trial court reached the "extraordinary" conclusion, as the appeals court put it, that the child had no legal parents. Reversing, the appeals court applied *Johnson* by analogy to conclude that husband and wife were the child's lawful parents.

As the *Baby M.* and *Johnson* cases illustrate, the law is much more receptive to gestational surrogacy than traditional surrogacy. This is true about legislatures as well as courts. Thus, for example, in North Dakota, traditional surrogacy agreements are void, but with gestational surrogacy, the intended parents are recognized as the child's parents, N.D. Cent. Code §§ 14–18–05 and 14–18–08. Similarly, more states permit gestational surrogacy than traditional surrogacy. Accordingly, the preferred

practice today when the intended mother cannot provide the egg is to use an anonymous egg donor and a gestational surrogate rather than a traditional surrogate.

A complete survey of the ethical and moral debates that still rage over surrogacy is beyond the scope of this book. The most willing

defense of the practice comes from those who view public policy issues economically. *See* Richard Epstein, *Surrogacy: The Case for Full Contractual Enforcement*, 81 Va.L.Rev. 2305 (1995); Richard Posner, *The Ethics and Economics of Enforcing Contracts of Surrogate Motherhood,* 5 J.Contemp.Health L. & Pol. 21 (1989). John Robertson, in CHILDREN OF CHOICE (1994), argues that the constitutional right of privacy should protect the right to enter and enforce surrogacy contracts, as a form of "non-coital" reproduction. Some argue against surrogacy contracts on the grounds that they are harmful to children. Margaret Brinig, *A Materialistic Approach to Surrogacy: A Comment on Richard Epstein's Surrogacy*, 81 Va.L.Rev. 2377 (1995). Other well-known commentators on this topic include Martha Field, SURROGATE MOTHERHOOD (1988) and Marjorie Shultz, *Reproductive Technology and Intent-Based Parenthood*, 1990 Wisc. L. Rev. 297, whose views were in part the basis of the California Supreme Court's analysis in *Calvert*.

Feminists have been of two minds. Some believe that allowing surrogate motherhood is consistent with ensuring all women full individual freedom of choice in making use of their own reproductive

capacities. Lori Andrews, *Surrogate Motherhood: The Challenge for Feminists,* 16 L. Med. & Health Care 78 (1988) (surrogacy is a "predictable outgrowth of the feminist movement"). Others worry about the commodification of women's reproductive capacities. E. Anderson, VALUE IN ETHICS AND ECONOMICS 168–190 (1993); Margaret J. Radin, CONTESTED COMMODITIES (1996). See also D. Callahan, *No Child Wants to Live in a Womb for Hire,* Nat'l

Catholic Reporter, October 11, 1985 ("Women object to being baby factories or sex objects because it offends their human dignity.").

One well-known ethicist has condemned surrogacy as baby-selling, the moral equivalent of slavery, George Annas, *Fairy Tales Surrogate Mothers Tell,* 16 L. Med. & Health Care 27 (1988). The Catholic Church has condemned surrogate motherhood on the basis of its belief that sex and reproduction cannot be uncoupled and cannot take place outside of marriage.

3. THE STATUS OF STORED EMBRYOS

There are a variety of medical conditions which leave a woman unable to become pregnant even though she has healthy eggs and a functioning uterus that would allow her to carry the child if she became pregnant. For example, a problem in the fallopian tubes may disrupt the normal movement of the eggs from the ovaries. In vitro fertilization (IVF) may be an effective fertility treatment in such cases. The woman's egg is removed, fertilized in vitro with her partner's sperm, allowed to mature for a few days, and then transferred to the woman's uterus with a

cervical catheter. If this embryo attaches to the uterine wall, a normal pregnancy may result. This procedure does not normally produce parentage issues, since the woman who bears the child is also its genetic and intended mother, and in the usual case, the genetic father is the mother's husband.

The difficulty of obtaining eggs for such a procedure, however, yields another potential problem. To spare the woman the experience of repeated procedures, IVF programs typically retrieve multiple eggs at one time, which may then all be fertilized. The

couple may have 10 or more embryos to use, but would want to use only 1–3 at first. The remaining embryos may be preserved in liquid nitrogen, available for implantation at a later date. But what if that later date never comes? In some cases, the couple becomes divorced before using all of the embryos. Legal principles make it clear that the couple decides about the disposition of the embryos, but what if the couple cannot agree?

In Davis v. Davis, 842 S.W.2d 588 (Tenn. 1992), a divorcing couple had seven frozen embryos in storage at a Knoxville fertility clinic where they had sought, unsuccessfully, to achieve a viable pregnancy. When the parties first sought divorce, the wife wished to continue her attempts to become pregnant through post-divorce implantation of the embryos, to which the husband objected. By the time the matter came to the state high court for review, both parties had remarried. The former Mrs. Davis now wanted to donate the embryos for use by another couple, while Mr. Davis had become quite firm in opposing any

procedure that would result in the birth of a child. The trial court viewed the embryos as "children in vitro" and treated the matter as a child custody dispute. It then awarded "custody" of the embryos to the wife. The Tennessee Supreme Court rejected this characterization of the issue and viewed an embryo not as a person but as human tissue deserving special respect because of its potential for becoming a child. The court then turned to a consideration of the interests of the Davises.

Mr. Davis strongly opposed fathering any child who would not live with both its parents, while his former wife's desire to donate the embryos was based on her desire to make some use of the

embryos after having gone through the lengthy IVF procedures. The court concluded that her interest in donation was less weighty than his in avoiding parenthood. Moreover, while agreeing that her interest would have been stronger had she sought to use the embryos herself, the court also suggested that it would still have been insufficient. Thus, while in form avoiding any bright line rules, the court's approach toward balancing would in most if not all cases seem to favor the party opposing use of the embryos. The court did leave open the possibility, however, that it might have decided this case in the wife's favor had use of these embryos been her only chance to bear her own children, particularly if she had not previously expressed a willingness to consider adoption.

In reaching its result, the *Davis* court noted that the parties had no prior agreement concerning the

disposition of the embryos. It indicated that if they had made such an agreement at the time they arranged for the IVF procedures, the agreement should be enforced (unless, of course, the parties later agreed to some other result). It appears that most IVF programs now routinely require participants to indicate on their consent forms how they wish to handle unused embryos. In Kass v. Kass, 696 N.E.2d 174 (N.Y. 1998), the consent form contained a section in which the parties could choose, from a provided list of alternatives, the disposition they preferred "[i]n the event that we . . . are unable to make a decision regarding the disposition of our stored, frozen pre-zygotes." The Kasses chose: "Our frozen pre-zygotes may be examined by the IVF Program for biological studies and be disposed of by the IVF Program for approved research investigation as determined by the IVF Program." When the parties later divorced, the wife wanted to attempt post-divorce

implantation with their five frozen pre-embryos, claiming that it was her best and perhaps last chance for genetic parenthood. Her former husband objected, and he prevailed when New York held that the executed consent form constituted an enforceable agreement between them on the matter which governed this dispute.

In Massachusetts, the state's highest court held that prior agreements are not enforceable, just as other pre-birth promises (e.g., a promise to place a child for adoption) are not enforceable. A.Z. v. B.Z., 725 N.E.2d 1051 (Mass. 2000). Other courts have stated that prior agreements are enforceable, but also have held that either member of the couple must

be allowed a change of heart before the agreement is implemented. In re Marriage of Witten, 672 N.W.2d 768, 782–783 (Iowa 2003); J.B. v. M.B., 783 A.2d 707 (N.J. 2001).

The possibility left open by the *Davis* court—that the woman might prevail if the frozen embryos offered her a last chance to have a genetically-related child—has materialized. In a few recent cases, courts have responded by awarding the embryos to the woman. In these cases, the embryos were created before the woman underwent chemotherapy for cancer that left her infertile. Terrell v. Torres, 438 P.3d 681 (Ariz. App. 2019) (on review in Ariz. S. Ct.); Szafranski v. Dunston, 34 N.E.3d 1132 (Ill. App. 2015); Reber v. Reiss, 42 A.3d 1131 (Pa. Super. 2012).

4. GENETIC MEDICINE

Modern genetic research allows medical science to move beyond merely facilitating the fact of conception, to altering, and

presumably improving, its product. No regular newspaper reader can be unaware of the frequent articles announcing the discovery of a gene for this or that disease, and perhaps many would not be surprised to read that scientists had found the gene for high LSAT scores or for playing a jazz trumpet. But genetics is more complex than that. Nearly every human trait of interest is affected by many different genes, as well as by the environment with which they interact. It matters little to be genetically lactose-intolerant in a culture that consumes no dairy products, and

aptitude for higher mathematics may not be discernable in a culture with no schools.

But while these cautions need be kept in mind, there is no denying that we are in the midst of an explosion of knowledge about human genetics that must have social as well as scientific implications. The Human Genome Project, an enormous research program involving scientists in dozens of locations, generated a complete map of the human genome. The next step, far more complex, lies in understanding the function of each gene. Many are of course already understood, and prenatal diagnosis for a number of genetic diseases is well-established. Genetic tests are also emerging that predict future health problems for adults, such as cancer or neurological disorders.

What shall we do with such knowledge? There are many applications and implications beyond the scope of this chapter, such as genetic screening by health and life insurers, and the possibilities for marketing laboratory-made genetic material and for claiming ownership of genetic processes and structures. Here, we describe briefly only the possible uses of genetic technologies that

are related to the reproduction issues surveyed in this chapter. We first consider the possibility that individuals may be offered the chance to learn whether they are afflicted with, or are carriers of, known genetic diseases.

Let us start with the fairly straightforward example of Huntington's Disease, sometimes known as Woody Guthrie disease. This neurological disease leads to impaired motor control, involuntary muscular contractions, personality changes, and

dementia. The defective gene is dominant, so there is no distinction between carrier and sufferer—a single gene, inherited from either parent, produces the affliction. The disease, while devastating, does not typically emerge until well into adulthood, giving the affected person plenty of time to bear children of his or her own—each with a 50% chance of inheriting the disease. So Alice might learn, at age 25, for example, that one of her parents has Huntington's Disease. Not that long ago, the result was uncertainty. Alice knew she had a 50% chance herself, but she might not yet show symptoms, even if she had the gene, since onset typically occurs between the ages of 30 and 50. Today genetic testing is available. Not every potential Huntington's sufferer has sought to be tested. Some prefer to remain uncertain rather than risk confirmation that they are affected. But then should such a person herself have children? If she is afflicted, half her children, on average, will be as well. Some believe it is not ethical to bear children knowing of their potential suffering—and of the possibility that their parent will not survive to their adulthood.

Of course, prenatal diagnosis is an option. Even if Alice is generally comfortable with having an abortion to avoid bearing a

child with serious genetic defects, is it a different matter for her to become pregnant knowing her heightened risk for that result? If so, is in vitro fertilization a choice? That is, Alice could have her eggs retrieved, have them fertilized in vitro, and then have them examined for presence of the Huntington's gene—and go forward

with the implantation using only embryos that lack the Huntington's gene.

Indeed, the ability to do pre-implantation genetic examination raises other possibilities. Some diseases are treatable with transplants from genetically compatible relatives. A lethal leukemia, for example, might be treatable with a stem cell transplant from a healthy sibling's bone marrow or blood. Of course, the sibling might not be a compatible donor. Some parents of children afflicted with an otherwise fatal illness have had additional children in the hope of producing a child who can serve as a life-saving donor for their sibling. Yet the chance of a child who is genetically compatible is less than 1 in 4. These parents, feeling under pressure to produce a healthy donor in time to save their existing child, might choose to become pregnant, obtain prenatal diagnosis, and then abort if the resulting baby could not be a donor, in order to quickly try again. But IVF technology makes this drastic and controversial measure unnecessary, for these parents now can instead test which of several embryos to implant in order to produce a healthy donor.

Techniques like these are limited only by the current state of genetic knowledge. At some point, scientists may be able to look at the embryos available for IVF and distinguish those with higher probabilities of being smart, tall, or blond from those with lower

probabilities for those traits. Should such knowledge be available for clinical application? If adults may choose to have purely cosmetic surgery, should they also be allowed to choose embryos whose

genes give their children an improved chance for maturing into more attractive adults? Needless to say, these questions do not now have legal answers, but at some point they probably will. Or, these issues may be forestalled by legal obstacles placed in the path of technology development that seems too threatening to social values. For example, the federal government refuses to fund research using fetal tissue because of the concern this might create a market for abortions. And the 1997 announcement by a British scientist that he had produced a genetic clone of a sheep named Dolly led to calls for banning or refusing to fund studies directed toward cloning humans.

It is sometimes thought that while it is possible to alter or control genetic makeup prior to birth, you are stuck with whatever genes you have when you are born. But this is no longer true. Research studies offer gene therapy—therapeutic interventions that alter the existing genetic structure in patients suffering various genetically affected diseases. These are still experimental protocols and there have been notable failures, but at some point they will lead to generally approved treatments that will be widely available. The issue raised in the preceding paragraph is no different in principle from those raised by gene therapy. Suppose gene therapy were available to improve the cognitive function of those congenitally afflicted with intellectual disabilities, such as sufferers of Down syndrome. If we are willing to modify the structure or functions of

their genes to improve their cognitive functioning, would we refuse to allow a different gene therapy that would improve

the cognitive functioning of those whose intelligence is in the "normal" range?

Finally, there is the possibility of preventive genetic medicine, through the use of genetic screening. Indeed, one might imagine mandatory genetic testing whose purpose is to inform individuals about important aspects of their own genes. For example, some people dramatically raise their chance of suffering a heart attack if they eat a high fat diet, while others are immune to such effects. Both would likely find it useful to know which they are.

Perhaps even more intriguing is the possibility of learning the particular genetic traits one might wish to seek or avoid in a mate. Many genetic diseases result from recessive genes. The person with only one abnormal gene in a gene pair may experience no important deleterious impact, because the normal other gene is dominant. But children who inherit the defective gene from each of their parents are seriously afflicted. It may seem advisable for people who share recessive genes for the same genetic disease to avoid having children together, for one of every four of their children will suffer the disease, and another two will (like them) be a carrier for it. Other alternatives for them include prenatal genetic tests, perhaps followed by abortion, or IVF in which embryos with two copies of the recessive gene are discarded.

The choice among these alternatives depends on one's values, but any of the choices is facilitated by knowledge that special risks are associated with

these two particular people having children together. The chance that two randomly chosen people from the population will share such deleterious defective genes is generally remote, but people do not choose their mates randomly from the population. They are indeed more likely to marry within their own ethnic group, and most ethnic groups have a heightened frequency of some genetic diseases (and a lowered frequency of others). For example, Tay-Sachs disease, while generally rare, occurs much more frequently among Jews of Eastern European ancestry than among the population at large, and sickle cell anemia is much more common among African-Americans.

Given the risks of genetic disease, an important question is when doctors should recommend testing to couples contemplating pregnancy. On one hand, couples who test positive for a genetic risk of disease can take steps to prevent transmission of the gene. On the other hand, testing can be expensive, and false positives may greatly exceed true positives when there is widespread testing for a rare gene.

Most feasible are programs targeted at specific high risk subgroups where testing is more reliable. Tightly knit subcultures in which marriages with outsiders are rare are particularly good candidates for community-wide programs that test for all genes for which that group is at high risk. And indeed, one such group in New York routinely tests children. Arranged marriages are the norm in this group. Confidential records are kept of children who are found in these routine tests to have a deleterious

recessive gene. When a marriage is proposed for that child, the intended mate is tested, and if the mate shares the trait, community leaders advise the proposed couple's families against the marriage, on the ground that the couple is "genetically incompatible." If such a program can work today in an appropriate subgroup, perhaps advancing technology will someday lower costs sufficiently to make it financially practical to propose more generally.

While genetic testing typically is reserved for individuals at elevated risk for an abnormal gene, there are important examples of testing that is recommended routinely. For example, under expert panel recommendations, doctors should offer cystic fibrosis testing to all couples contemplating pregnancy, in order to inform partners who both have this recessive trait that they are at 1-in-4 risk of having affected children. Routine testing had been controversial because of the costs relative to the low incidence of the disease in the general population (about 1 in 1000) and because the test had a significant false negative rate (that is, it missed a large number of actual carriers because not all of the mutations were known). With improvements in testing, routine screening has become more justifiable. Down syndrome is another disorder for which routine screening has replaced targeted screening. While pregnant women once were offered testing if they were at least 35 years old or had a family history of Down syndrome, experts now believe that physicians should offer testing to all pregnant women.

These issues are complex and at this point often speculative, and so we only touch upon them here. While it may be possible to respond adequately to a number of these issues with existing legal rules, many of these advances in genetic technology are sure to

pose fundamental challenges and expose serious weaknesses in the existing legal structure.

C. THE RIGHT NOT TO PROCREATE

Just as the Constitution protects the right to choose procreation, it protects the right not to procreate. The government cannot prohibit access to contraception, and it cannot prohibit access to abortion.

One can view rights of access to contraception and abortion as permitting the decoupling of sex and reproduction. Without access to contraception or abortion, having sex can readily lead to becoming a parent. The right not to procreate is also a right to sexual liberty, a right recognized more directly by the Supreme Court in Lawrence v. Texas, 539 U.S. 558 (2003) (striking down a ban on oral or anal sexual intercourse between two persons of the same sex).

1. CONTRACEPTION AND ABORTION

a. Contraception

In 1965, the Supreme Court took its first step to recognizing a right to contraception by striking down a Connecticut ban on the use of drugs or devices to prevent pregnancy. Griswold v. Connecticut, 381 U.S. 479 (1965). The Court was troubled by the

practical implications of the law—it effectively empowered the police to invade the privacy of the marital bedroom. While the *Griswold* decision tied the right to contraception to marriage, later cases defined the right contraception as a right of all persons, married or single, adult or minor. The Court also recognized a right not only to use contraceptives but also to purchase them.

In *Griswold*, the Justices invoked multiple constitutional provisions for a right to contraception that was rooted in the protection of personal privacy. In later cases involving abortion, the Court tied the right to privacy in reproductive decisions to the due process clauses of the Fifth and Fourteenth Amendments, ultimately recharacterizing the right as a liberty right. Planned Parenthood v. Casey, 505 U.S. 833 (1992).

b. Abortion

Once the Court had recognized a right not to procreate by using contraception to *prevent* pregnancy, logic suggested a corollary right not to procreate by using abortion to *terminate* pregnancy. In 1973, the Court recognized such a right in Roe v. Wade, 410 U.S. 113 (1973). According to the *Roe* Court, the right to privacy encompassed the right to terminate a pregnancy because of the risks to women and their families:

> Specific and direct harm medically diagnosable even in early pregnancy may be involved. Maternity, or additional offspring, may force upon the woman a distressful life and

> future. Psychological harm may be imminent. Mental and physical health may be taxed by child care. There is also the distress, for all concerned, associated with the unwanted child, and there is the problem of bringing a child into a family already unable, psychologically and otherwise, to care for it. In other cases, as in this one, the additional difficulties and continuing stigma of unwed motherhood may be involved. Id at 153.

As with other rights, the right to abortion is not absolute. According to *Roe*, regulations limiting the right to abortion needed to satisfy the traditional "strict scrutiny" standard for fundamental rights—the regulations had to promote "compelling" state interests and be "narrowly drawn" to serve those interests.

And the state clearly has very strong interests to justify regulation. For those who believe that life begins at conception, limits on abortion serve one of the most important state interests, that of preserving life. In addition, just as childbirth comes with risks to the woman's health, so does abortion.

In applying the strict scrutiny standard of review to abortion, the *Roe* Court established its famous "trimester" framework. During the first trimester of pregnancy, the government could not regulate abortion because the state's interests in fetal life and maternal health, while important, were not yet

compelling.[3] By the second trimester, the risks to maternal health from abortion were significant enough to make the state's interest compelling and therefore a basis for regulating abortion. Once the fetus became "viable" (which at the time of *Roe* roughly coincided with the end of the second trimester), the state's interest in potential human life was compelling, and the government could prohibit abortion. Still, even after viability, abortion had to be allowed if continuing the pregnancy would threaten maternal health.

Over the subsequent two decades, the Court heard multiple challenges to *Roe*, and in the 1992 *Casey* case, the Supreme Court modified its *Roe* framework, replacing strict scrutiny with an "undue burden" standard. According to the Court, the government may regulate abortion throughout pregnancy to promote its

important interests in maternal health and potential human life, as long as it does not impose an undue burden on the right to choose abortion. That is, the government may not erect a "substantial obstacle" to abortion access. And as in *Roe*, the government may prohibit abortion after viability, except when abortion is necessary to protect maternal health. Planned Parenthood v. Casey, 505 U.S. at 878–879.

421

(1) The Undue Burden Standard

In elaborating on the undue burden standard, the *Casey* Court expanded state regulatory authority in a significant way. To promote its interest in potential life, the government may adopt measures throughout pregnancy "to ensure that the woman's choice is informed." In addition, it is permissible for the government to try "to persuade the woman to choose childbirth over abortion." Id. at 878. Accordingly, while in previous cases the Court had invoked *Roe* to strike down waiting periods or mandatory disclosures of information by physicians to abortion patients, the *Casey* Court approved these kinds of regulations.

In *Casey*, for example, the Court upheld a requirement that doctors disclose not only the nature of the abortion procedure, the health risks of abortion and of childbirth, and the probable gestational age of the fetus. The physician or a qualified aide also had to inform the woman of the availability of material that described the fetus and supplied information about medical assistance for childbirth, child support from the father, and agencies that provide adoption and other services as alternatives to abortion. Id. at 881–885. *Casey* also upheld a requirement for a 24-

hour waiting period between the woman being informed about abortion and the abortion being performed. Id. at 885–887.

Other permissible regulations according to *Casey* and other cases include bans on funding for abortions that are not necessary to save the woman's life and requirements for minors to notify, or secure

permission from, a parent, as long as the minor can avoid parental involvement through a judicial hearing.

The Court also upheld a federal ban on an abortion procedure in which an intact fetus is removed from the woman through the birth canal ("partial-birth" abortion). Gonzales v. Carhart, 550 U.S. 124 (2007). Typically, after the early second trimester, abortions are performed via a procedure in which the cervix is dilated and the fetus removed in pieces ("dilation and evacuation" or "D&E"). For later second trimester abortions, some physicians used a D&E procedure that allowed for the removal of the fetus in one piece ("intact D&E" or dilation and extraction, "D&X"). That procedure became controversial because a large part of the fetal body was pulled through the cervix before fetal life was terminated. Critics of intact D&E and the *Carhart* Court likened the procedure to infanticide.

While the Court has upheld some regulations of abortion, it also has invalidated other requirements, such as (a) for the woman to notify the would-be father of her plans to have an abortion, (b) for the physician to have admitting privileges at a nearby hospital in case there are complications from the abortion (rather than relying on physicians staffing the hospital to provide care), or (c) for the clinic providing abortions to satisfy the safety standards for outpatient surgical centers.

Spousal notification requirements are not allowed because of the risk to women from abusive partners. The Court struck down the admitting privileges and

safety standard mandates on the grounds that they did not actually make abortion safer while at the same time substantially impeding access to abortion—the difficulties in meeting the requirements resulted in widespread closures of clinics providing abortions.

It is difficult to be very certain about the implications of the undue burden standard. Given its inherent imprecision—when exactly does a burden create a "substantial" obstacle to abortion—judges and Justices can easily come to different conclusions. Thus, for example, while the Supreme Court struck down the admitting privileges law in Texas, Whole Woman's Health v. Hellerstedt, 136 S. Ct. 2292 (2016), it did so only by a 5–4 vote, and a panel of the U.S. Court of Appeals for the Fifth Circuit subsequently upheld a Louisiana admitting privileges requirement by a 2–1 vote. The Supreme Court currently is considering the constitutionality of the Louisiana statute in June Medical Services v. Gee.

The Court has exacerbated the uncertainty of the undue burden standard by defining it in different ways from one decision to another. In *Casey*, for example, the Court seemed to describe a two-part test—did the government have an "important" interest for its regulation, and if so, did the regulation promote that interest without creating a substantial obstacle to abortion?

But in the partial-birth abortion case, the Court relaxed the *Casey* standard in a few ways. First, the Court permitted the government to invoke a

"legitimate" rather than "important" interest. More specifically, at stake was a symbolic, moral interest—the intact D&E procedure that was banned looked like infanticide. But the law only changed how an abortion would be performed rather than whether it would be performed and therefore did not promote the state's interest in preserving potential human life. In addition, the ban did not promote the state's interest in maternal health. Indeed, the district court had concluded that banning the intact D&E procedure actually would increase the risk to health for some women.

This reflected another relaxation of the *Casey* standard of review —the Court called for deference to the legislative judgment on whether prohibiting the procedure would increase the risk to maternal health. In contrast to the district court, Congress concluded that the intact D&E procedure did not reduce the risk of an abortion in some cases. Hence, the Court upheld the ban even though it lacked an exception for cases in which the procedure would make abortion safer for the woman.

Subsequently, in *Whole Woman's Health*, the Court tightened up the undue burden standard, rejecting the idea that courts should defer to legislative assessments of health risks. But the Court still deviated from *Casey* in another way. Under *Casey*, abortion regulations may not create a substantial obstacle to abortion. Under *Whole Woman's Health*, though, the Court described the undue burden standard as a balancing test, in which the burdens of a law must be considered together

with its benefits. Id. at 2309–2310. Rather than simply asking whether or not a regulation imposed a substantial obstacle to abortion, the *Whole Women's Health* Court asked whether the benefits of the regulations were great enough to justify the burdens they imposed on pregnant women seeking an abortion.

Because many states have adopted a range of new regulations of abortion, the Supreme Court will likely refine the undue burden standard further. In addition, with retirements and appointments of new Justices, it is possible that the Court will give states broader discretion to regulate than it has to date.

(2) Viability

Drawing a line at viability for a right to abortion has been criticized by both sides of the debate. On one hand, it is argued, if we determine the end of life by the cessation of cardiac or neurologic function, we should determine the beginning of life by the initiation of cardiac or neurologic function, both of which occur before viability. On the other hand, it is argued, pregnant women should be able to control what happens to their body, including whether to maintain a fetus inside of it.

Despite the criticisms, a line at viability can be justified for important reasons. As the *Casey* Court observed, viability "is the time at which there is a realistic possibility of maintaining and nourishing a life outside the womb, so that the independent existence of the second life can in reason and all

fairness be the object of state protection that now overrides the rights of the woman." 505 US. at 870.

In addition, consider an argument made by Judith Jarvis Thomson (*A Defense of Abortion*) and others. Instead of viewing abortion as a killing, we can view it as a decision by the woman to no longer provide life support for the fetus. In other words, rather than thinking of abortion as like euthanasia, we can think about it as like the withdrawal of life-sustaining treatment.

Ordinarily, under U.S. law, people have obligations not to harm others, but they do not have obligations to provide assistance. The classic example in tort law is the absence of a duty to come to the aid of someone drowning. If the law imposes a duty to carry a pregnancy to term, then it would single out women for a duty to aid that no one else has. In other words, a right to abortion reflects principles of equal protection.

Of course, the lack of a duty to aid is qualified in special relationships, such as the parent-child relationship. People do assume duties of care to their children. But even then, legal duties stop short of aid that would impose a risk to the parent's health. So, for example, while parents must provide food and shelter to their children, they do not violate the law if they refuse to donate a kidney or stem cells from bone marrow or blood to save their child's life. Carrying a pregnancy poses risks to the woman's health and therefore goes beyond the level of duty that the law imposes on parents.

D. MATERNAL-FETAL CONFLICT

Abortion is the most prominent maternal-fetal conflict considered by courts and legislatures, but it is not the only one. From conception to childbirth, the lives of the mother and her unborn child have a biological and psychological connection that is unique.

Medicine traditionally focused its attentions on the mother, having little it could offer to improve the child's prospects other than keeping the mother in good health. The unborn child was not accessible to the doctor, and thus could not be treated as a separate patient with its own needs. But over the past several decades, this has become less true, for two reasons. First, we have much more information today about the effect on fetal health of particular maternal behaviors such as smoking, drinking, or using drugs. Second, physicians have a much greater ability to detect fetal distress and therefore recommend treatments for the mother that will help the fetus. A new ethical puzzle has thereby been created: to what extent, if any, is the mother ethically or legally obligated to follow medical advice reasonably thought necessary to protect the health of her child?

This section divides maternal-fetal conflict cases into two categories, depending upon whether the pregnant woman is refusing an unwanted treatment or is engaging in behavior harmful to her fetus. In the first category, many of the cases involve refusals of blood transfusions or cesarean sections. In the second category, many of the cases involve use of illicit drugs. In the future, we may see a different type of

treatment refusal—women declining surgery on the fetus in utero, a process which of course also requires surgery on the mother.

An obvious starting point to analyze these maternal-fetal conflict cases is abortion law. It is tempting to conclude that if a woman can abort her fetus before viability, she has no duties of care to the fetus before viability. Similarly, if a woman cannot abort her fetus after viability, she assumes duties of care to her fetus upon its viability.

As we will see, though, there are important distinctions between an abortion, on one hand, and a refusal of treatment or use of harmful substances on the other hand. Hence, principles of abortion law do not necessarily carry over.

For example, there is no necessary inconsistency in combining a public policy that recognizes a woman's right to terminate her pregnancy with another that imposes obligations on her with respect to the conduct of her pregnancy *if* she chooses to continue it to term. The law recognizes a duty of parents to promote the health of their children. Since harm to child health can occur before birth, we might extend the duty to promote child health to pregnant women. Or to put it another way, the "greater right" to abort a fetus doesn't necessarily include the "lesser right" to harm a fetus that will be born.

That said, there will be ways to promote the future child's health that would entail a more onerous burden than that which the law would require of a woman with regard to her born children. In those

cases we might reject a duty of care for the pregnant woman even if we believe that she owes some duty of care to her fetus for lesser burdens.

It is not only the case that a right to abortion doesn't imply the absence of a duty to promote fetal health when childbirth is planned, it also is the case that the lack of a right to abortion after viability doesn't imply that the woman has a broad duty to promote fetal health after viability. Abortion entails a deliberate action harmful to the fetus while a refusal of a blood transfusion or a

cesarean section is an omission. While some omissions are as blameworthy as acts, many are not.

1. FORCED TREATMENT AND FETAL SURGERY

Forced treatment cases often have involved blood transfusions or cesarean sections, and for both kinds of cases, courts have come to different conclusions about the pregnant woman's obligations to her fetus.

At work in these cases are a number of considerations. For example, one might take the view that the right to refuse life-sustaining treatment should not be limited by virtue of a person's pregnancy. Otherwise, women would be forced to choose between preserving their right to refuse unwanted treatment and exercising their right to reproduce. In this view, one should not have to waive the right to refuse treatment as a condition of having children.

A good example of a court taking this view is an intermediate court of appeals in Illinois. The court held that blood transfusions could not be ordered over the objection of a Jehovah's Witness, then about 35 weeks' pregnant, despite medical evidence that transfusions were the only available treatment that would raise her hemoglobin levels. The doctors testified that if her hemoglobin levels were not raised, both she and her baby had only a 5% chance of survival. In re Brown, 689 N.E.2d 397 (Ill. App. 1997). The *Brown* court rejected any balancing of the woman's and fetus' interests, concluding that pregnant women enjoy the same right to refuse unwanted medical treatment as other persons.[4]

Another important principle lies in the prohibition against risking one person's health for the benefit of another. For example, as

discussed, even after fetal viability, states must permit abortions when necessary to protect the woman's health. Consider also the example of living kidney donation. Because most people have two functioning kidneys, and a person can do nearly as well with one kidney as with two, living kidney donation is encouraged. Nevertheless, because there are risks to the donor's health, the law does not require kidney donation.

This no-trade-off principle is reflected in the tragic case of McFall v. Shimp, 10 Pa.D.C.3d 90 (1978), in which the plaintiff was fatally ill with aplastic anemia. His only hope was a stem cell transplant, and Shimp, a cousin, was the only member of his

family whose bone marrow was sufficiently compatible to be a donor. Shimp refused. Bone marrow extraction is a painful process, and it carries risks, but not serious ones. McFall sued, seeking to compel the life-saving donation from his cousin, but the court turned him down. Although it found Shimp's conduct morally reprehensible, the court concluded that to require him to donate marrow "would change every concept and principle upon which our society is founded. . . . For a society which respects the rights of one individual, to sink its teeth into the . . . neck of one of its members and suck from it sustenance for another member, is revolting to our hard-wrought concepts of jurisprudence."

This line drawn at bodily invasion is a familiar one. For example, it is well established, even in the criminal law, that one cannot compel witnesses to undergo surgery or, for example, to have their stomach pumped, in order to provide access to critical evidence.

Of course a parent's obligation to his or her child is greater than a cousin's, and so we can surely require more of a mother than a

cousin. But there is no precedent for compelling an unwilling parent to assume a risk to health for a child. Such cases would of course be rare—parents and siblings are generally willing to donate stem cells or even kidneys to a family member in need, just as pregnant women are generally willing to accept a blood transfusion or deliver by cesarean section where that is necessary for their child. The closest case involving a parent is probably Application of George, 630 S.W.2d 614 (Mo.

App. 1982), involving an adult adoptee who needed medical information from his biological parents in order to identify compatible stem cell donors essential to the treatment of his leukemia. A blood test of his father was critical. The petitioner's father was identified in the confidential adoption records, and he was located by the trial judge. He refused to cooperate, denying that he was in fact the father. The trial judge went to great lengths to persuade him to submit to a blood test; it was clear that the judge believed strongly that this man should cooperate. But when he refused no order was entered and his identity was not disclosed to the applicant.

The no trade-off principle is particularly relevant to the cesarean section cases. Their occurrence reflects in part the sharp increase in the cesarean section rate. The proportion of all childbirths in the U.S. that are by cesarean increased from about five percent in the mid 1960's to about 32 percent in 2007, where it has remained since. While the cesarean-section rate is much too high, there are many times when it reduces the risk to the child. In some of those cases, a cesarean section will promote the health of both mother and child, but in other cases, it increases the risk to maternal health (with a mortality rate of about two in 100,000 deliveries).

In re A.C., 573 A.2d 1235 (D.C. 1990), is illustrative. In that case, a pregnant woman, A.C., experienced an untreatable recurrence of cancer during her 25th week of pregnancy. While she initially stated her desire to have the fetus delivered by cesarean section at week 28, her health

deteriorated quickly, and her mental status became uncertain. According to medical testimony, an immediate cesarean section would give A.C.'s fetus a 50–60 percent chance of survival, but it might shorten A.C.'s life (which was expected to last no longer than another 24–48 hours). After the trial court ordered a cesarean section, the D.C. Court of Appeals held that pregnant women's wishes should be respected absent "extraordinary or compelling reasons" to override her decision. In discussing a decision by the Georgia Supreme Court permitting a forced cesarean section (discussed below), the D.C. court observed that in the Georgia case, "the evidence showed that performance of the caesarean was in the medical interests of both the mother and the fetus." 573 A.2d at 1243.

An Illinois court of appeals came to a similar conclusion as *A.C.* In Baby Boy Doe, 632 N.E.2d 326 (Ill.App. 1994), a woman was diagnosed at 35 weeks of pregnancy with a placenta defect that compromised the oxygen supply to her baby. Doctors recommended immediate delivery by cesarean section, which the woman rejected on the basis of her "personal religious beliefs." The County Attorney filed a wardship petition, but the court sustained the mother. It did not rely on her vague religious objection, but rather held more generally that a competent woman's choice to refuse a medical procedure "as invasive as a cesarean section" must be honored, even if the choice will harm her fetus.

The court rejected any balancing of the mother's interests against the fetus', and distinguished earlier cases involving blood transfusions on the ground that, as

compared to cesarean sections, such transfusions are "relatively non-invasive and risk-free." (As mentioned above, the later Illinois decision in *Brown* rejected the distinction between a blood transfusion and a cesarean section in permitting a pregnant woman to refuse an unwanted transfusion.)

If the no-trade-off principle precludes mandating treatments that pose risks to the woman's health, what if the treatment would benefit maternal, as well as fetal, health? One could take the view that it is permissible to impose a duty on parent to child or pregnant woman to fetus when there is no risk to the parent or pregnant woman's health. Some courts have taken that view, particularly in blood transfusion cases.

For example, in Application of Jamaica Hospital, 491 N.Y.S.2d 898 (1985), a woman was bleeding heavily during the 18th week of her pregnancy, and doctors testified that the woman and the fetus were at high risk of death without a transfusion. The court ordered the transfusion, on the ground that the state's interest in protecting fetal health outweighed the woman's right to refuse unwanted treatment. While the court didn't provide much discussion, it and other courts probably have been willing to order transfusions not only because they are not particularly invasive treatments but also because the transfusions serve the health interests of both woman and fetus.

The Georgia Supreme Court authorized doctors to perform a Caesarean section over the objection of a pregnant woman, if the

surgery was medically

necessary to protect the health of the woman and her 39-week-old fetus. Jefferson v. Griffin Spalding County Hospital Authority, 274 S.E.2d 457 (Ga.1981). In that case, the physicians were concerned that a complete placenta previa jeopardized maternal and fetal health. While placenta previas typically resolve naturally before labor, if they do not, the health risk is high—according to the trial testimony, a vaginal delivery would have a 99% mortality rate for the fetus and a 50% mortality rate for the woman. If the placenta previa did not resolve by the time labor commenced, a cesarean section was permissible. See also Pemberton v. Tallahassee Memorial Regional Medical Center, 66 F. Supp. 2d 1247 (N.D. Fla. 1999) (ordering a cesarean section to promote both fetal and maternal health).

Similarly, recall that the *A.C.* court did not reject forced cesarean sections in all cases, leaving open the possibility of a court order in extraordinary or compelling circumstances. In that regard, the court specifically declined to approve or disapprove its decision in a previous case requiring a cesarean section. In commenting on that case, the *A.C.* court observed that "there was no real conflict between the interests of mother and fetus; on the contrary, there was strong evidence that the proposed caesarean would be beneficial to both. 573 A.2d at 1252 n.23.

As indicated, three principles tend to drive the cases in this area: (1) pregnant women, like other persons have a right to refuse unwanted medical treatment, (2) if pregnant women choose to carry

their fetuses to term, they have some obligation to ensure good health for their children, and (3) one person's health should not be sacrificed for the benefit of another person's health. In reconciling these principles, courts generally will not impose treatment on a pregnant woman unless it provides health benefits for her as well as for her fetus, and some courts will not impose treatment even when it provides a health benefit to the pregnant woman.

There are other additional concerns to consider. Critics of an approach based on maternal autonomy observe that the no-trade-off principle should not apply in the context of pregnancy. Unlike McFall's need for his cousin's stem cells, or a child's need for a parent's stem cells, the unborn child is fully dependent on the mother's decision; no one else can help the fetus.

Critics of limits on maternal autonomy, even when treatment would benefit the woman's health, also cite important factors. For example, the great majority of women involved in the cases are poor minority mothers, and a large proportion are unmarried and do not speak English. Rights to make medical decisions belong to all persons.

In addition, many who work in hospital obstetric departments believe that in nearly all cases in which consent is not initially forthcoming, it can be obtained through better communication with the pregnant woman. What is usually needed is not coercion, but personnel sensitive to the woman's cultural traditions and fluent in her native language, able to successfully counsel her to consent. In the rare

cases in which the patient persists in declining treatment, it's not clear that the law can provide an acceptable remedy. Suppose the woman defies the court order. Should medical personnel be authorized to restrain her forcibly to carry out the procedures over her objection? Most find that prospect distasteful, at the least.

It is nonetheless not unprecedented. In *Brown*, *supra*, the trial court had ordered the contested transfusion, which was carried out before the appellate decision reversing the trial court's order. The court reported that the doctors "yelled at and forcibly restrained, overpowered and sedated" the objecting patient. In *Baby Boy Doe, supra*, the public guardian and physician, apparently unwilling to attempt a cesarean section under such conditions, did not actually seek an order authorizing the operation over the mother's objections. They instead asked the court to provide that she could be held in contempt if the baby was born either dead or severely intellectually disabled as a result of her refusal. Consider whether the threat of contempt is likely to alter the behavior of a woman who has not responded to a careful explanation of why her decision itself poses these very risks for her child.

The difficulty of designing an acceptable remedy in the cesarean cases is accompanied by a concern that legal coercion may ultimately harm more children than it helps. This concern was one reason for the policy statements of the American College of Obstetricians and Gynecologists and the American Medical Association. In discouraging legal action,

ACOG and the AMA considered not only the pregnant woman's immediate autonomy claims, but also the more general impact of such legal actions on medical practice. By resorting to legal action

to enforce their advice, physicians may not only risk destroying their relationship with that particular patient, they also may discourage other pregnant women from seeking medical assistance. More babies overall might therefore suffer from the effort to save one child.

2. FETALLY TOXIC MATERNAL BEHAVIOR DURING PREGNANCY

While there are good arguments against compelling a woman to undergo surgery for the benefit of her developing child, it seems as if the government would be on stronger ground if it forbids smoking, alcohol or drug use, or an unhealthy diet. How should we assess such regulation? And what if the government required the woman's partner to avoid smoking in her presence? On one hand, these kinds of requirements impose no bodily intrusion and no risk to personal health. On the other hand, they do require unusually intrusive regulation of personal conduct that in some cases is otherwise lawful.

Arguably, the most appropriate legal tool for such intrusion is the civil child abuse and neglect law. It is intended to protect children, and the ultimate remedy ordinarily is removal of the child from the parents' custody, a step often taken when the child cannot otherwise be protected. It is often hoped that parents will alter their conduct to avoid such

removal. Direct orders regulating parental conduct are less common. But when the fetus is our concern, the threat of removal may prove less helpful. Removal after the child's birth is in principle available, and has sometimes been used, as many states treat conduct during pregnancy to be within the scope of child

protective statutes. But see Pima Cty. Juvenile Severance Action S-120171, 905 P.2d 555 (Ariz.App. 1995) (fetal abuse cannot be the basis for termination of parental rights).

Post-birth removal has its limits, however. It cannot reverse the harm done to the child during the pregnancy, and thus does not address the problem unless it serves to deter the pregnant woman from behavior toxic to the fetus. The potential for post-birth removal may deter some, but clearly not all. Is any other remedy possible? Can one effectively regulate maternal behavior during pregnancy? Can one incarcerate the recalcitrant pregnant woman to compel her compliance? Should one? These are the questions we address here.

There is surely no doubt that the problem is serious. Consider alcohol abuse. In a study of first-grade children in a representative Midwestern community, researchers found a rate of developmental disorders from prenatal exposure to alcohol of 2.4 to 4.8 percent. 134 Pediatrics 855 (2014). In communities with high rates of alcohol abuse, the harm to children is even greater. Prenatal exposure to illicit drugs also hinders childhood development, though not as much as does exposure to alcohol.

The law's experience with maternal drug use is not new. For many years, some states have ruled that the child of a narcotic addict is per se an abused child, and therefore allow removal of the child from the mother at birth. Others hold specifically that harm caused a child during pregnancy by the mother's addiction can be the basis of a neglect finding. In some states, standards for the termination of parental rights have been expanded to specifically include drug dependency which the parent has failed to have

treated successfully. These laws sometimes distinguish criminal drug abuse from non-criminal behavior posing similar risks on children. For example, the New York Social Services Department, in an administrative proceeding, held that a child abuse report cannot be established by evidence that the mother drank excessively. In re D.W., 10 Fam. L. Rep. 1359 (B.N.A.1984). But not always. See Fla. Stat. Ann. § 39.806(1)(k) (2019), treating drugs and alcohol alike (presence of alcohol or a controlled substance in the blood or urine of a newborn infant can be grounds for termination of parental rights).

But of course a rule that drug abuse during pregnancy may support a finding of child neglect after birth does not address the irreversible damage that those undeterred by the rule will cause their child during pregnancy. In some cases, of course, pregnant substance abusers who seek help may find no treatment program available to them. But in other cases, the problem is more complicated. Consider, e.g., Angela M.W. v. Kruzicki, 561 N.W.2d 729 (Wis. 1997). Angela's obstetrician suspected her of cocaine abuse. He confronted her with blood test findings and

counseled her to enroll in an inpatient drug treatment program. Her response was to cease coming for pre-natal care. He then reported her to county authorities, who relied upon child abuse statutes to obtain a court order allowing them to take "custody of the fetus"—and thus custody of Angela—if she did not remain in a treatment facility. The order was reversed by a divided Wisconsin Supreme Court, which held that Angela's fetus was not a "child" within the meaning of the relevant state statute. They did not rule on whether such a remedy would be constitutional if the legislature provided it.

Efforts to address the issue through the criminal law rather than the civil child abuse laws have also met with limited success. One must observe that the policy goal of these prosecutions is unclear. The cases are generally brought after harm has already occurred, so that no remedial purpose is possible, unless one believes some general deterrence will be achieved. The Supreme Court of Florida reversed the conviction of Jennifer Johnson, whose child was born with traces of cocaine in its blood. Johnson v. State, 602 So.2d 1288 (Fla. 1992). The prosecutor argued unsuccessfully that Ms. Johnson violated the criminal law by "delivering" cocaine to her child via the umbilical cord, during the 1½ minutes between the emergence of the child's head from the birth canal (after which it was a "child" and not a "fetus") and the clamping of the cord. The Court held the statute was not intended by the legislature to apply to such facts. Prosecutions in other states brought on the *Johnson* theory have met the same fate.

442

So have prosecutions based on a theory of child abuse. Courts typically have held that child abuse statutes don't apply to fetuses because they are not yet children. A good example is the case of Pamela Rae Monson who was charged with violating a statute making it a misdemeanor to "willfully omit . . . to furnish necessary" medical care to one's child. The prosecution ended in a dismissal, on the grounds that the statute had not been intended to apply to prenatal acts or omissions. Ms. Monson was charged with ignoring her physician's advice to stop using amphetamines and marijuana for the remainder of her pregnancy and to seek immediate medical assistance if she began to hemorrhage. Her child, born with massive brain damage, died at six weeks. The principal exception to this pattern is Whitner v. State, 492 S.E.2d

777 (S.C. 1997), which sustained the application of a criminal child endangerment statute in South Carolina to a woman's ingestion of crack cocaine during the third trimester of her pregnancy, holding that a viable fetus is a "person" within the meaning of that statute. Tennessee passed a fetal endangerment statute in 2014, but it lapsed after a few years.

Should pregnant drug abusers be placed in some form of involuntary confinement if necessary, to protect their fetus—hopefully, in well-run treatment centers? On the practical level the principal objection is that such a program would be ineffective, or even harmful, because the affected women, like Angela M.W., would in consequence avoid pre-natal care, and otherwise try to keep themselves from coming to the authorities' attention. That argument is plausible

but hard to evaluate. If real effort and resources were put into enforcement, and into creating proper placements for those affected, it might perhaps succeed.

There are of course potential constitutional issues. Consider People v. Pointer, 199 Cal.Rptr. 357 (App.1984). The defendant mother was convicted of criminal child endangerment as a consequence of her fanatical insistence that her two small children adhere exclusively to a macrobiotic diet that had brought one of them near death. She remained unrelenting throughout, even endangering the sicker one by covertly breast feeding him in the hospital, in violation of the physician's explicit instructions (her own diet made her milk dangerous to the child in his weakened state). She fled with her children to Puerto Rico at one point, in an ultimately unsuccessful effort to evade the authorities and continue

her children on the disputed diet. Faced with these facts and others that persuaded him that the defendant would endanger any future children as well, the trial judge imposed, as a condition of her probation, a requirement that she not conceive. A sympathetic court of appeals reluctantly overruled this order, concluding that while its purpose was "salutary," it nonetheless unconstitutionally burdened the defendant's privacy rights by effectively requiring her to obtain an abortion if she became pregnant, or have her probation revoked.

The court's concern about abortion exists for any woman who is subject to criminal sanctions during pregnancy, as, for example, under the child

endangerment statute applied in *Whitner*. If a woman addicted to cocaine or other illicit drugs becomes pregnant in South Carolina, she can avoid prosecution only by aborting her fetus. Or to put it another way, drug-addicted women in South Carolina lose their right to procreate by virtue of their addiction. This is even more troubling when one considers that according to the National Institute of Drug Abuse, "drug addiction is a disease of the human brain." Nevertheless, the *Whitner* court took the view that if drug use is illegal, there are no constitutional barriers to punishing its use during pregnancy.

Even the *Pointer* appeals court didn't fully recognize the constitutional problems. That court urged, as an alternative and presumably acceptable condition to a restriction on procreation, that Pointer be tested periodically for pregnancy and that if she became pregnant, she be required to follow an intensive "prenatal and neonatal treatment program monitored by both the probation officer and by a supervising physician." It then remanded to the

trial court to allow it to fashion such an alternative set of conditions. The appeals court in *Pointer* thus offered, as the acceptable alternative, the very kind of pregnancy regulation that many would question as excessively intrusive. In *Pointer*, such monitoring could presumably be imposed as a condition of probation. Would it be less burdensome on the defendant's privacy than simply requiring her to remain nonpregnant? Would it be any more practical to enforce?

445

These and related questions may seem particularly relevant in considering how to deal with drug-abusing pregnant women. Some report that trial judges often use their sentencing discretion to incarcerate pregnant women convicted of drug abuse—or sometime, of other crimes, if they are also drug users—for a period at least long enough to bring them to their due date, even if others might receive lesser sentences or even probation. Involuntary assignment to a drug treatment program, as was attempted in *Angela M.W.,* seems preferable, and certainly more straightforward. It is perhaps ironic that while constitutional issues might be more easily avoided if pregnant drug abusers are confined pursuant to the criminal rather than civil law, the civil law may be both a more effective and more humane tool.

There is nonetheless the legitimate fear that civil regulation initially aimed at behavior constituting criminal drug abuse might end up casting a wider net. It could easily extend beyond controlled substances to abuse of alcohol, an extension many would surely urge. But then, is confinement also appropriate to suppress occasional alcohol consumption, or chain smoking? Few would go that far, but what can be offered as the limiting principle? One relevant consideration is the state of the scientific evidence.

Suppose available evidence supports claims that heavy drinking during pregnancy has a high probability of causing specifically identified problems, while the harm risked by smoking is less well-identified, and less likely to occur.

Scientific evidence might also bear upon the intrusion necessary to forestall the harm. For example, as neuroscientists learn more about the development of the brain in utero, they may be able to pinpoint specific windows of fetal vulnerability to exposure to particular toxic agents. Suppose, for example, we learned that the damage caused by alcohol abuse could be avoided almost entirely if no alcohol is consumed between the 13th and 16th weeks of pregnancy? Perhaps regulation pinpointed at this vulnerable period could be justified where more wide-ranging regulation would not be.

For the most part such possibilities lie in a future world with more certain scientific information. But the questions can be put most clearly by examining one particular example that has received attention in the literature, which presents a useful paradigm for testing our beliefs on these questions.

PKU is a congenital enzyme deficiency rendering the infant incapable of metabolizing phenylalanine. At one time children born with this defect became severely disabled intellectually, but this result is now avoided through a mass screening program on all newborns. Affected infants are placed on special diets. The traditional medical understanding has been that they need not continue the diet after reaching physical maturity, because their continued inability to metabolize phenylalanine causes no difficulties in adults. (More recent scientific research casts some

doubt on this traditional view, suggesting the diet may aid the cognitive functioning of PKU adults as well, but for purpose of this discussion, we

may ignore this development.) In recent years, women born with PKU who benefitted from neonatal screening programs have now reached adulthood and become pregnant themselves. These now-pregnant PKU adults must resume the limited diet of their childhood, not for their benefit, but to avoid disastrous consequences for their baby. While the unmetabolized phenylalanine in the mother causes her little harm, it apparently crosses the placenta in such amounts that it overwhelms even a normal baby's capacity to deal with it. The result is severe intellectual disability, an abnormally small head, congenital heart disease, and other difficulties. There is no effective treatment after birth for most of the baby's problems.

There is, however, a simple, effective, and risk-free program that allows the PKU mother to shield her baby from this fate. She must, during pregnancy, go back on the low phenylalanine diet she ate as a child. She can eat some normal foods, in carefully controlled amounts—breads, cereals, fruits, vegetables. But her main source of protein is a specially prepared paste of L-amino acids to which vitamins, minerals, carbohydrates, and fats have been added. This paste tastes bad, and the mother's access to normal foods is restricted. But if she keeps to this diet during her pregnancy her child has the same prospect of health as other children. If she does not, her child will suffer irremediable, and possibly catastrophic, abnormalities. If the mother were to refuse to comply with her physician's advice to adhere to this diet, should a judge be empowered to order her to comply?

Some commentators argue such direct coercive sanctions can be justified only if the intrusion is minimal and would avoid a severe handicap in the child that physicians were very confident would otherwise occur. An example might be case in which the needed treatment requires only a one-time intervention of minimal risk to the mother, such as a single administration of a drug of known safety. Effective intervention with the PKU mother, however, would involve long-term confinement and control over her diet. Moreover, it seems unlikely that allowing this kind of remedy could in fact help the child of the recalcitrant PKU mother, since by the time one would have the evidence necessary to establish the need for intervention, she would have already doomed her child by departing from the necessary diet.

Other writers see any efforts to intervene in the pregnancy as but another example of gender-based restrictions on women's autonomy. Certainly, any policy allowing increased impositions on a pregnant woman's autonomy to aid her unborn child must be part of a larger, gender-neutral policy that reaches men as well. For instance, men (and women) might be compelled as parents to donate blood, or possibly stem cells, to their born children if necessary to save a child's life or prevent substantial harm. Thus, even though pregnancy is unique to women (or transgender men), a principle that would allow us to sometimes intervene in pregnancy could apply to men as well. But this would cast in doubt cases like *McFall*, which involved only a one time intervention

of stem cell donation, not long term coercive confinement.

For now these questions have no clear legal answer. It seems likely, however, that in future years, new medical knowledge will allow them to be put in more compelling forms.

[1] Today, sterilization sometimes can be reversed.

[2] In one difference compared to *K.M.H.*, the donor and the mother had lived together earlier and tried to have children through sexual intercourse.

[3] While the *Roe* Court wrote that regulation was not permitted during the first trimester, it did make room for some regulation, such as a requirement that abortion be performed by a physician.

[4] A trial court had permitted the transfusions, so the appellate court's decision governed future cases.

INDEX

References are to Pages