

Cost Containment and the Quality of Medical Care: Rationing Strategies in an Era of Constrained Resources

DAVID MECHANIC

Rutgers University

BOTH PHYSICIANS AND THE GENERAL PUBLIC BELIEVE that cost-control efforts are possible without impairing the quality of medical care. In recent surveys commissioned by the American Medical Association, 71 percent of doctors and 86 percent of the population interviewed agreed that "medical care costs can be reduced without reductions in the quality of care" (Freshnock 1984). After two decades of vigorous economic expansion and technical development in medicine—with great public support—increasing proportions of the public and especially public policy makers give more priority to other areas of national life. Between 1978 and 1984 health care dropped in public priority rankings from first to sixth place, and now significantly trails behind financial support for the elderly and education. More efforts will be made to hold the growth of public programs in health and tax subsidies to the private sector for expanded medical care benefits within controllable limits. This need is especially pressing in an atmosphere of rapid evolution of new sophisticated biomedical knowledge and technology, the aging of the population, and the growth of the physician pool and other health personnel.

The wisdom of imposing significant constraints on a dynamic industry undergoing rapid scientific and technological development, however,

is not obvious. Despite its concerns, the public continues to value highly the advancement of biomedical knowledge and technology and comprehensive entitlements for medical care coverage. But the realities of cost escalation, and its consequences for the competitive position of American business, government budgets, and other competing sectors of the economy, will require us to choose between more forceful limitations, significantly higher taxes, or increased patient cost sharing. The proportion of national income which ought to be invested in health care is a value judgment and a product of the political process, but there is wide skepticism on whether or not the marginal benefits of additional medical care for the average consumer either deserve or require investments much beyond present proportions. Such notions, of course, are changeable, and sufficiently impressive medical advances could conceivably build support for much larger contributions of the gross national product. This appears unlikely, however, and cost-containment efforts are expected to be an important focus for the foreseeable future.

The source of concern about cost relates less to aggregate expenditures and more to the tax burden of public-sector programs and the public subsidies to nonprofit and proprietary health endeavors. When faced with competing claims on national resources, government finds it easier to restrain growth in programs affecting the poor and disabled, who constitute relatively weak constituencies, than to reduce subsidies shared by large, articulate, and sophisticated segments of the larger American public. The scope and mode of financing in the health sector overall has increased the cost within public programs of meeting the needs of those with the most sickness but the least personal resources. The imminent risk we face is not a deterioration in medical care overall, but more a continuing erosion of access and appropriate care for our most unfortunate populations. The poorest populations, particularly those who depend on Medicaid, are most vulnerable to loss of access and limitations on scope of services during times of economic stringency. Between 1976 and 1984 the proportion of poor and near poor covered by the Medicaid program decreased from 65 to 52 percent. New initiatives in cost containment that do not achieve balance in relation to the entire system of care will result in different and inferior levels of care for the poor in contrast to the overall population. Evidence for this trend is already becoming apparent (Lurie et al. 1984).

The medical care arena is presently in ferment. Federal and state governments are exercising more regulatory muscle as evidenced by rate regulation and the introduction of diagnostic related group methodologies and professional review organizations. The hospital sector is evidencing a strong, corporate proprietary interest, the diversification and unbundling of services among nonprofit voluntary hospitals including spinoffs of private corporations, and mergers and increased collaborative activities. Interest in competition and competitive incentives have helped health maintenance organizations (HMOs), preferred provider organizations, and a variety of free-standing ambulatory facilities for surgery, emergency care, home care, etc. The increase in physician supply has provided opportunities for innovation and altered practice styles that would have appeared highly unlikely just a decade ago. Between 1981 and 1984 the proportion of physicians who reported that they would consider joining an HMO increased from 25 to 33 percent (Freshnock 1984). Just where this mix of public and private interests, and competition and regulation, will lead remains unclear but the dynamic quality of the industry suggests major changes and innovative future options.

Despite the evident dynamism of health services activities, attitudes are very much polarized. At one extreme are physician leaders who view the system as fundamentally sound, who strongly support traditional fee-for-service payment schemes and who oppose such alternatives as HMOs (Iglehart 1984a). At the opposite pole are critics who view the delivery system as so flawed in its structure and priorities and so dominated by special interests that only major reorganization offers any promise of an equitable and effective delivery system in the future (Sidel and Sidel 1983). Given the range of viewpoints and interests, it is inevitable that a diversity of models will persist, reflecting the vast differences among local situations and necessary compromises with powerful groups having a major stake in public policies.

There is speculation that the size and power of profit-oriented multihospital corporations will reduce the traditional influence and autonomy of physicians (Starr 1982), particularly in the context of an impending "physician excess." The aggregate numbers may appear alarming in light of physicians' potential to generate costs well beyond their own income as managers of patient care, but the growth of physician manpower also provides opportunities for hospitals and other medical care institutions to innovate in practice organization and care

patterns in a way that doctors could easily resist when they were in short supply. Such physician dominance, for example, not only slowed the growth of prepaid group practice but also insured that managers of such organizations could not very forcefully intervene in how resources were being used (Friedson 1975). Despite the changes noted, hospitals continue to rely on physician cooperation to maintain bed occupancy in an increasingly competitive arena and, thus, continue to satisfy physician demands for technology and other expensive, but not always essential, facilities. Bringing physicians under the diagnosis-related group (DRG) hospital rate could alter both physician behavior and the hospital's willingness to satisfy desires concerning technology and facilities.

Consideration of Some Workable Constraints

The growth of third-party insurance, and the willingness to pay providers on the basis of costs or usual and customary fees, provided incentive for increased provision of services relative to available capacity. Under such payment schemes, the greater the supply of physicians, hospital beds, and ancillary capacity, the greater the number of services provided. While there are practical limits to the ability of providers to utilize excess capacity, the greater the available resources the more likely they are to be used. Altering third-party insurance by increasing coinsurance and deductibles and imposing limits on types of coverage inhibits such provider-generated demand, but there is sufficient clinical uncertainty even under more stringent payment conditions to allow significant variations in medical decision making (Wennberg, McPherson, and Caper 1984).

It is now commonplace to argue that increased competition among providers will reduce costs but distinctions, unfortunately, are rarely made between unit costs and aggregate costs for populations. While the increased number and diversity of providers may set limits on what a specific service may cost, aggressive competition can generate new types of services, new demands, increased uncoupling of services, repeated hospitalizations and visits, and the use of more ancillary care. In many sectors of the economy, growth in aggregate expenditures reflects how individuals decide to invest their income at the margin. Thus, if airfares are less expensive and more people travel, both the industry and consumers may benefit. In contrast, if competition in

health care results in lower unit costs but more aggregate utilization, the results, depending on the procedures and technologies used, could be injurious to health. No one would contend that more surgery is "good" unless clearly linked to improved health benefits.

Few physicians consciously exploit patients or cynically manipulate the reimbursement system. Uncertainty, however, provides a context in which specialty bias, personal inclination, and economic interests are easily confused with quality care and appropriate practice. In uncertain situations, physicians are more likely to practice the skills they know and feel comfortable with, and that yield economic rewards.

Rationing: General Considerations

There are basically three approaches to constraining costs, and all in some sense are rationing devices (Mechanic 1979). These are: significantly increasing cost sharing both to deter utilization and share the burden of paying for those services that are used; establishing capitated systems that set general budgetary limits on providers but also allow these systems to establish internally their own priorities and expenditure patterns; and formally regulating health care costs through explicit decisions on coverage, acquisition of facilities and technologies, and conditions for service provision. Most countries, including the United States, use some mix of all three approaches but differ significantly in how they balance them.

Although rationing is sometimes evoked by critics of change as a new impending threat and aberration, rationing of health care has always existed but the ways it has been achieved are sufficiently consistent with common modes of thinking to attract little attention. Rationing is no more than a means of apportioning, through some method of allowance, some limited good or service. Given the complexity and generosity of the American system of care, we have successfully maintained the illusion that rationing is foreign to it.

Rationing by Cost Sharing

When individuals paid directly for their own care out of pocket, decisions were always influenced implicitly by cost and competing economic needs by both patients and their caregivers. As insurance

covered more costs, such considerations became less influential but persisted. Even now few persons have insurance with full coverage for all services without deductibles and coinsurance, and such cost sharing is significantly increasing. It is estimated that major employers requiring a hospital deductible more than doubled between 1982 and 1984. The Rand Health Experiment demonstrated that coinsurance can have a powerful inhibitory effect (Newhouse et al. 1981). Persons receiving free care had expenditures about 50 percent more than those with income-related catastrophic insurance.

While there is little evidence that increased cost sharing, associated with reduced ambulatory care, adversely affects health in the aggregate, there is some evidence that the poor do less well under cost-sharing conditions (Brook et al. 1983). In many instances, cost sharing is likely to inhibit the poor and old more than those who are affluent, yet these are the populations with the greatest medical problems requiring the most care. Moreover, coinsurance arrangements are administratively complex, and it is expensive to administer them in a way that provides incentives for providers to be more cost conscious and efficient. Since coinsurance and deductibles are unpopular among consumers who value comprehensive coverage, and are easily manipulated by public programs during periods of economic stress, they also contribute to uncertain patient expectations and instability in the system of care.

Explicit Rationing

Cost sharing has been creeping up in such important programs as Medicare, and in many third-party insurance programs in the private sector, but other regulatory efforts have accelerated as well. Such efforts can conceptually be differentiated in terms of the extent to which they seek to restrain costs by overall budgetary limitations without clinical intrusion, and whether they more explicitly seek to regulate expenditures for varying components of care and the adoption of new technologies. In the latter instance, government health agencies explicitly control the acquisition and diffusion of new technologies and procedures, specify the range of care, types of service, context of care for reimbursement, identify reimbursable providers, and even specify the frequency or minimal intervals for carrying out varying tests, procedures, and examinations. While there are many instances of each of these types of regulation already in such public programs

as Medicare and Medicaid, and in private and nonprofit health insurance programs as well, such explicit limitations have most typically involved areas that were outside the conventionally defined core of covered services. Thus, limitations on service coverage are more likely to exist for mental health, dental and drug utilization, and new "experimental" procedures than for "basic" components of general medical care. Certificate-of-need regulations were designed to control the spread of expensive new technologies but were not applied to existing ones, nor did they much influence such acquisitions outside hospitals. In short, government has trod carefully in decision making involving clinical authority. If costs are to be seriously controlled by explicit means, more forceful intrusion on physician autonomy and clinical work is necessary, but it is likely to meet vigorous resistance. At its best, explicit rationing is informed by technology assessment, health services research, clinical investigation, and careful assessment of costs, benefits, and tradeoffs. But it is inevitably tainted by political pressures that affect government action, and administrators making decisions are too often distanced from the pain, anguish, and uncertainty of serious illness and the responsibilities for its management.

Implicit Rationing

Alternatively, constraints can be applied implicitly by establishing general limits as reflected in a budget with specified levels of growth. Implicit approaches allow the transactions between institutions, health professionals, patients, and families to determine the potential payoffs and tradeoffs. In theory, implicit rationing frees professional decision making from the tyranny of fee incentives and the excesses and distortions associated with them, allowing patients' needs to define the type, intensity, and range of services to be provided, and professionals to weigh necessary tradeoffs. Those advocating such budgeting approaches have argued that provision of service in such systems is determined by need and not by demand. The advantage to administrative authority is that it attains its basic purpose of setting overall constraints but frees it from involving itself in the complex details and uncertainties of clinical practice.

The obvious question is the extent to which theory fits reality and to what degree the suggested advantages of freeing clinical decision making from fee incentives are realized. One would anticipate that

administrators organizing services with fixed budgets would seek to avoid providing unnecessary services, would substitute preventive and psychosocial services for medical services when appropriate and economical, and would use nonmedical professionals such as nurse practitioners, nurse midwives, nurse specialists, social workers, psychologists, and others in expanded roles wherever possible. The evidence is reasonably persuasive that capitated practices achieve economies, particularly by reducing hospital admissions and total hospital days. But performance in other areas approximates medical practice more generally.

Health Maintenance Organizations

In the case of capitated practice, we in the United States have had more experience with HMOs than any other alternative. Having grown in number and market penetration only very slowly since the 1970s, more recent federal and private-sector encouragement and support has resulted in accelerated expansion. At first count there were almost 17 million enrollees in HMOs and enrollments have been growing yearly at 15 to 20 percent. In some areas they have become major contenders for a significant market share. The potential is illustrated by developments in Minneapolis-St. Paul where 6 nonprofit plans enroll 36 percent of the population (Iglehart 1984b).

HMOs are a good model of implicit rationing in that once the capitation is determined, the organization can establish its own priorities, modes of service delivery, mix of professional personnel, balance of services among prevention, acute care, and chronic disease management, and many other matters. For many years the literature indicated that prepaid group practice, particularly the large established ones such as Kaiser Permanente, achieved considerable cost savings by reducing hospital admissions by as much as 10 to 40 percent (Luft 1981). While many glib explanations were advanced, the most likely explanation was the limitation on available hospital beds and the absence of economic incentives for hospital admission. Some critics suggested, however, that the effects might be explained by the types of persons who elect to join HMOs.

In the Rand Health Experiment, 1,149 persons were randomly assigned to an established HMO, the Group Health Cooperative of Puget Sound, and an additional 733 prior enrollees were also studied

(Manning et al. 1984). Both groups had 40 percent fewer hospital admissions than in the free fee-for-service plan, a finding that supports the aggregate literature on the cost effectiveness of HMOs.

Many studies have noted that HMOs provide more preventive services than office-based fee-for-service physicians, but Luft (1981) has argued that this difference is related more to the comprehensiveness of insurance coverage characterizing HMOs than to any unique aspect of its organization or incentives. In contrast, the Rand Health Experiment found that preventive visits in the HMO studied were twice as frequent in comparison to the various fee-for-service plans, and even exceeded those made in the free fee-for-service plan. Most such preventive services were for well-child care and gynecologic examinations. Out-patient visits, in contrast, did not significantly differ between the HMO and the free fee-for-service plan.

HMO advocates have often maintained that cost savings result from the incentive to keep people healthy, treat them early, and avoid subsequent need for more intensive services such as hospitalization. There has never been any significant support for this contention and all indications point to the lower hospital admission rate as a result of the style of practice rather than differential morbidity (Mechanic 1976, 83–98). Even in the ambulatory care situation, HMOs have not consistently provided more access than other insurance programs, although the HMO studied by Rand provided more outpatient visits than any but the free fee-for-service plan. For example, while those randomized into the HMO made 4.3 visits per person, those in the experimental group with 25 percent coinsurance used 3.5 visits. There is little overall evidence from the experiment, or from other studies, that these variations in access to outpatient care have the significant impacts on health that some believe them to have.

In appraising the experience and future potential of HMOs, it is important to note that HMOs emerged within a skeptical and often hostile environment and, until recently, faced significant difficulties in recruiting physicians. In many communities, both the organizations and the physicians associated with them were “on trial” and were harassed by organized medicine. Thus, they followed a conservative course both to facilitate physician recruitment and to avoid controversy and adverse publicity. Past failure to be innovative in many of the ideal terms suggested by advocates may say little about future potentialities in the supportive and encouraging environment that is

now developing. The difficulty of recruiting physicians is no longer an acute issue and the antipathy found among many older physicians toward HMOs is less common among more recent physician cohorts.

Until now, medical services in HMOs have been dominated by a physician perspective, one which mirrored medical practice in fee-for-service practice, partly for the reasons already suggested. The little evidence that is available suggests that the physician's style of practice and use of practice modalities in HMOs reflects prior training and clinical orientations more than organizational management. In the future, the supply of physicians, and their greater willingness to work in HMOs, will allow HMO administrators more leverage in supervising physician behavior and in innovating a mix of services. Although we have not, as yet, seen a great deal of innovation, the potential is there. Other health practitioners will have to demonstrate, of course, that their substitution for increasingly available physician services is, in fact, cost-effective and that the expanded services they offer in preventive, psychosocial care and health maintenance help limit demand for other components of the health care package.

Rationing Styles: Areas of Uncertainty

The theory underlying the presumed effects of removal of fee incentives on professionals also requires scrutiny. The evidence is that removal of such incentives decreases the motivation and work effort of physicians to some degree, allowing them to substitute more leisure for a longer work week. There is little cause for alarm in the context of a growing supply of physicians and other health care personnel as long as these professionals remain responsive to patients. There is little evidence of difference in work orientations by type of practice, but the frequent finding that patients in HMOs feel their physicians less interested in and responsive to them than in office-based, fee-for-service practice merits some watching (Luft 1981).

In thinking about these issues for the future, the distinction between how an organization is paid and how doctors are paid is fundamental. HMOs, despite their status as capitated systems, may choose to pay physicians and other personnel in varying ways depending on the quality of their motivation and work, their responsiveness to patients, and their productivity. Capitation insures that the basic income pool

is fixed; how such income is distributed to best effect is an area deserving of more attention than it has heretofore received.

In HMOs, where physician remuneration is unrelated to the number or type of procedures they perform, it is naive to expect that only medical need will now determine the allocation of effort; nor can we assume that professionals left to their own devices will have as their only priority the interests of their patients. It is human to have preferences, to pursue the interesting and novel rather than the dull and routine, and to seek activities and practices held in higher esteem by one's peers and the world at large. It does not follow that physicians, health administrators, nurses, and other professionals will allocate resources in direct relationship to need as compared with the attractiveness and sophistication of the patient, the inherent challenge and excitement of varying procedures, career needs, and personal inclinations. Patients, too, vary greatly in what they demand, their degree of acquiescence, and their skills in manipulating the system, and there is little in human experience that should lead us to believe that the transactions that result will bring comparable services or outcomes. Variability is to be expected; what must particularly be guarded against is allocation of disproportionate resources to the more affluent and sophisticated patient with less medical need but higher expectations and greater persuasiveness. This area needs careful attention, particularly as more Medicaid and Medicare recipients are enrolled in HMOs.

Experience also teaches us that equal access is a goal more easily talked about than achieved. All systems of care are under great pressure from educated and sophisticated constituencies that know what is possible and increasingly demand it. While our system of health services is sufficiently generous to soften the consequences of differential social and political pressure, it is also likely that the system will yield when persistently pushed. Thus, the groups at greatest risk are the sick, poor, old, and less educated clients who neither know the system nor how to manipulate it. Standards to protect such vulnerable populations should be a high priority.

A budget is simply an instrument; the scope and quality of care, and the types of rationing, depend, of course, on how it is managed. Budgets require health administrators and professionals to set priorities and it is here that we have the least information and face the greatest uncertainty. While, in theory, professionals will carefully examine the tradeoffs and constrain their colleagues within some reasonably

established priorities and criteria, there is little evidence that this occurs. Administratively, it is easier to delay the initiation of a new technology, service, or unit, reduce staff and close beds, eliminate "nonessential" services, and constrain wages and other major costs. Only rarely, and at the margins, have efforts been made to alter significantly processes of care and decisions physicians typically make. Organizations and professionals must more directly confront variations in practice and cost that cannot be justified on the basis of differences in patient mix, the particular populations served, and the uncertainties of clinical efforts (Wennberg and Gittelsohn 1982).

The ability to ration services without divisiveness depends on the public's trust. In the United States such trust has been high but fragile because of the divisions within the health professions and the increasing visibility of such differences in the mass media. Given the public's high expectations and support for medical technology, it is inevitable that the media will give attention to rationing efforts and their economic, ethical, and human consequences. Decision making in the spotlight is always more difficult and potentially divisive. American patients are less deferent to authority and more questioning than those in Great Britain, for example. Thus, we should not anticipate comparable acquiescence to limitation of resources and withholding of services that typifies the British context (Aaron and Schwartz 1984). The British have learned to accept many limitations in services and social amenities and much of the older population still recognizes the improvements in access to care that followed the establishment of the National Health Service (NHS). The stable and continuing relationships most people have with a general practitioner, limited geographic mobility, and the free availability of care at the time of service, all help reinforce the authority and trust placed in the general practitioner. As Aaron and Schwartz (1984, 104) note with respect to the unavailability of chronic dialysis to those over 55 years of age:

For many patients with renal failure, the local physician does not even raise the possibility of dialysis. In other circumstances, however, he says that dialysis does not seem to be indicated. Because of the respect that most patients have for physicians, the doctor's recommendations are usually followed with little complaint.

Another important dimension that facilitates rationing decisions, as Aaron and Schwartz note, is the stability of professional relationships

within the NHS and the common interest among both general practitioners and consultants to maintain effective continuing communication. In the United States the competing interests of physicians are greater, relationships are more fragmented, and doctors are more likely to compete openly and criticize one another.

It is fair to suggest that the degree of authority and trust that characterizes the provision of medical services in Britain exceeds any reasonable expectations we can have about the United States. Trust is also likely to erode in Britain as the population becomes more sophisticated and earlier generations are replaced by later cohorts who have no memories of services in the era preceding the NHS. Trust has always been an indispensable aspect of effective patient/physician encounters, and as physicians take on more responsibility for balancing their efforts as agent of the patient with the need to use resources in a cost-effective way, trust becomes even more essential. Indeed, one argument for rationing by external authority is that it insulates the doctor/patient relationship from the tensions of adjudicating between patient needs and demands and the budgetary constraints on the organization (Fried 1975). Under explicit rationing, physicians can more easily explain limitations on the basis of external constraints in contrast to their own decisions, and can be patient advocates without reservations.

The increasing tendency to "lock" patients to health provider organizations and limit choice subtly shifts the physician's role from agent of the patient to a bureaucratic official allocating resources among competing demands (Mechanic 1984). These changes create new tensions that, in the absence of trust-building procedures such as mechanisms for review and resolving patient complaints, could be troublesome (Mechanic 1978).

The Social Content of Rationing Decisions

Capitated systems need not be based on values different from fee-for-service organizations, but they have more incentive to avoid expensive institutional care and technical services. As we provide care for larger numbers of persons who are old, we will have to rethink carefully as a society what we consider appropriate care. The population aged 85 years and older increased 174 percent between 1960 and 1980, and

is expected to increase another 110 percent by the year 2000 (Rice and Feldman 1983). The potentialities for medical care for such populations are almost limitless, and such expenditures must be measured in value against other needed social and long-term services for this population. In 1980, 11 percent of the elderly in America accounted for 31 percent of personal health expenditures (U.S. Department of Health and Human Services 1984, 146–47). As the elderly population itself includes greater proportions of very old persons, this population will require even more of total health care resources.

Medical care, however, is only one component of the health care and functioning needs of the aging population. We have yet to face up to financing the long-term care of the elderly in a rational way. The finite resources available make it certain that we will face tough decisions on the relative value of varying types of services. Capitation lends itself to making such judgments and one alternative being examined is the Social Health Maintenance Organization (SHMO) (Hamm, Kickham, and Cutler 1982). This approach attempts to replace the fragmentation of financing and categorical services with a single system responsible for acute hospital and nursing services, ambulatory medical care, and such personal services as homemaker, home health and chore services, meals, counseling, transportation, etc. In the SHMO the provider organization takes on all financial, programmatic, and care decision responsibility within a capitation restraint. On Lok, a successful demonstration of the concept in San Francisco, illustrates the potentialities of such an approach (On Lok Senior Health Services 1983). SHMOs have many financial uncertainties but also great potential. They build on the widespread and deep desire of most elderly to remain in their homes and communities and avoid institutional care. To be financially viable, however, enrollment must be limited to elderly eligible for institutionalization or other costly care alternatives, and not significantly opened to new populations. The dilemma is that the more attractive long-term care services become, the more eligible clients are likely to appear.

Politics and Rationing

In the British National Health Service, rationing is a product of the unavailability of resources and a more or less implicit understanding among physicians about the appropriate limits of care. There is no

mandate that says that the elderly will not receive hemodialysis, but such patients are not referred for such care, and it is generally understood that they are not appropriate candidates. Such rationing through consensus reduces political debate over social policy issues in medical care as compared with explicit rationing by central authority, but it puts a special obligation on professionals to administer such understandings in a fair way. In the NHS, as in western systems of national health insurance more generally, it is rare for administrators or even fellow physicians to interfere with clinical judgment.

A major symbolic value of medical care is the effort to ensure all citizens equal opportunity to develop and use their capacities consistent with their aspirations. An effective system links in a value sense all classes, races, regions, and age groups. The imposition of political judgment as a substitute for professional discretion threatens not only the care process but also the symbolic value of the health care system. Some totalitarian societies structure health services to prefer strategic workers or government cadres over others. Obvious political rationing occurs in the United States as well when the Medicaid program denies payment for effective services such as abortion, generally available in other health care programs. Effort in the United States to impose criteria that arise from motives other than those concerned with quality of health and functioning could have damaging consequences for public trust and confidence. Without such trust, the system of care will inevitably deteriorate.

Rationing and Competitive Markets

However we organize future medical services, it is inevitable that public programs such as Medicare and Medicaid, and many employer-sponsored programs, will have limits on the scope of insurance. Patients will have options varying from relatively basic to more comprehensive benefits involving alternative deductible and coinsurance arrangements or additional premiums depending on the levels of coverage sought. The popularity of Medigap policies among the elderly attests to the desire for comprehensive services available at the point of need with limited or no out-of-pocket expenditures.

In Britain, supplementary private insurance markets reduce pressures for expanded or enriched services, dampen the frustrations and impatience of persons with high expectations, and also establish a competing

standard of service by which the mainstream system can be appraised (Klein 1983). By expediting services for those who wish to bypass the queue, such insurance also makes a modest contribution to reducing expenditures. In the United States, in contrast, we can anticipate relatively aggressive competition for enrollees in mainstream private sector and nonprofit health insurance programs. The emergence of new organizations offering selected services for urgent care, convenience, home visits, and other supplementary needs suggests that there will also be considerable competition in care provision at the margins. It is difficult to conceive, however, given the public's expectations and the value placed on medical service, that the basic care package can exclude any generally recognized and accepted medical modality. This is already apparent in the HMO sector where despite uncertain evidence of efficacy, HMOs have generally followed the pattern of community practice in the use of coronary and other forms of intensive care. Even when administrators have had serious doubts about the cost effectiveness of such units, the acceptance of them both among physicians and by the public, and the fear of malpractice litigation, has encouraged the HMO to follow the pattern of community practice.

It seems most likely that rationing will occur primarily in the area of amenities, in the intensity of diagnostic and laboratory investigation, and in the discretionary use of hospitals and surgical interventions. It is less likely, contrary to British experience, that rationing will occur by withholding or significantly slowing the use of new technologies of demonstrated effectiveness, particularly in disease areas of major visibility to the public like cancer and heart disease. There are significant cost savings to be achieved and the course need not be too painful given the apparent willingness of the public to see an increase in the medical share of the national income.

The trends suggest that we can expect no major transformation in the manner in which medical work is carried out or in the configuration of dominant providers. HMOs will undoubtedly capture more of the market, and profit-oriented hospitals and health care plans will play an important role, but we can anticipate a considerable mix of alternatives. We should expect more effective competition among insurance programs, greater care and judiciousness in the use of expensive procedures and technologies, and more control over marginal services and amenities. The system as a whole is more likely to evolve by muddling through and by individual groups taking advantage of new market

opportunities and incentives, than by broad efforts to rationalize and reform the system.

Persistent Uncertainties

The evolutionary process in health care has the strength of avoiding major disruptions and dislocations and leaves the majority of the population with confidence that adequate access and quality of care will be available to them. A minority of the population, and that segment that is most vulnerable and most needs protection, the poor and chronically ill, are likely to face the greatest threats since these populations depend almost exclusively on government programs that the public feels ambivalent about. The elderly are an increasingly powerful lobby, with widespread political support, and while we may see some changes in cost sharing, premium structures, age of eligibility, and the like, it is difficult to anticipate that the fundamental core of the program, which is of great attractiveness to a broad population base, will be dismantled. This is in sharp contrast to the Medicaid program which has from its inception been associated with welfare and the antipathy of the population to the "undeserving poor." What is typically not appreciated is that Medicaid is substantially allocated to care for the poor elderly whose Medicare benefits are insufficient, particularly for long-term care not covered by Medicare. With economic recession and increased fiscal pressures on states, Medicaid programs have failed to grow in relation to need, eligibility has been increasingly restricted, and many new regulatory and reimbursement constraints have been introduced that limit services or make recipients less attractive to hospitals, physicians, and other providers. When the going gets tough, it is this population that has the least countervailing power and public support, and it is here that persistent vigilance will be required. While better use of available resources may be possible with changing concepts of providing effective long-term care, the future of this entire arena, and its potential for overwhelming the public sector, remains an issue not yet adequately confronted.

A second uncertainty is the degree to which new conditions change the way physicians and other health providers behave. The introduction of DRGs under Medicare is simply a foot in the door; by itself it is unlikely to change fundamentally the way doctors or hospitals behave.

If expanded to include physician services and all payers, it has greater potential, although ultimately cost will depend as much on the politics of pricing as on the structural system in place. While capitation tied to individuals rather than diagnostic episodes would eliminate propensities to disaggregate services, increase admissions, and engage in other bureaucratic gamesmanship, a comprehensive diagnosis-related group (DRG) system would facilitate competition among competing health insurance plans. Both budget constraints and the growing number of physicians and other health professionals provide significant opportunities to modify decision-making processes. More physicians will work for organizations, or under some capitated arrangement, involving changed incentives. While underservice becomes a risk with a capitated approach, Americans are sufficiently demanding and sensitive to be vigilant. The altered power relationships between insurance programs, health organizations, and physicians should make doctors somewhat more cooperative, and the awareness of operating within a zero-sum budget may induce greater physician responsibility for over-seeing cost-effective patterns of care.

Structural modifications do not ensure changes in physicians' traditional resistance to administrative authority. Doctors are still trained to give priority to their clinical experience in making judgments of appropriate patient management. They typically are pragmatic and action-oriented, and are not particularly committed to theory, conclusions based on aggregate data divorced from clinical experience, or to administrative authority (Friedson 1970). Doctors are socialized to see themselves responsible for the individual patient, and this combined with the strong value of the primacy of clinical judgment makes them resistant to administrative authority, even when medically based. Such independence makes it difficult for either administrators or other physicians to exercise control over medical work, particularly in loosely organized medical settings.

Physicians will continue to be reluctant to sanction peers, a characteristic of professionals more generally. Given doctors' desires to preserve discretion for themselves, it is not surprising that significant control is difficult to achieve. It is unrealistic to anticipate that physicians will tightly control their colleagues nor is it particularly desirable. But incentives are possible that encourage group pressures to constrain behavior that cannot be justified by reasonable appeals to uncertainty or needs for discretion. One important place to start is in the area

of large variations in procedures performed that cannot be explained by the morbidity of populations served nor justified on the basis of improved outcomes. If physicians are unwilling to address the implications of such variations, it is inevitable that others must do so.

One of the most important uncertainties is how future public expectations evolve and how they come to affect the political process. At present, the public recognizes cost problems but expresses less concern than administrators, legislators, health experts, and industry and union executives. The public puts high value on the potentialities and contributions of the doctor, wants the best available technologies, and supports research efforts on major disease problems. It wants more rather than less medical care coverage, welcomes the growth of supply, and values immediate and responsive access. Since the vast majority of medical care costs are paid indirectly, and not by the patient at the point of service, it is difficult for patients to see the relationship between what they pay for medical care and the possible tradeoffs relative to other valued products and services. One of the most persistent findings in opinion polls is that the public supports increased coverage even if it means higher taxes. Moreover, many persons individually are willing to pay the price for more comprehensive coverage so that they can make care decisions for themselves and their families in a noneconomic context. System designers and consumers may have different objectives.

How public perceptions emerge in the future with changing demographic conditions and mounting long-term care costs will importantly affect the impending debate. People clearly recognize the cost pressures at an aggregate level but they do not want constraints when they or their loved ones need care. Demand for care can be constrained by public policies that reduce tax exemptions associated with health insurance benefits or, alternatively, employers could reduce their contributions to health fringe benefits that presently encourage high utilization. But it is equally clear that employees do not readily give up health benefits already attained through collective bargaining and there is strong resistance to changes in tax policy affecting health insurance. The American public, which has exceedingly high expectations of the medical care system, will not easily be persuaded that their welfare lies in a high degree of cost sharing or in delaying the use of promising new technologies. The surveys, however, suggest that the public is willing to cooperate to some degree in solving the

problem of cost escalation. Considerable ingenuity will be required in offering options that provide new and attractive outpatient alternatives as a tradeoff for accepting plans with a gatekeeper in relation to institutional and other, more expensive services. HMOs are presently the most attractive available alternative but considerable marketing skill and other incentives will be necessary to enroll large segments of the population.

One evident point is that, in the future, medical care cannot make available all that science and technology can potentially contribute. As in every other area of our economic life we will be faced with choices and tradeoffs and these will in no way be easy or uncontroversial decisions. At present, large expenditures relative to the population are being made at the beginning and end of life. The wide adoption of neonatal intensive-care technology and its application to allow survival of babies of lower and lower birth weights results in a large imbalance between what we spend to save a life and what we invest in that child's health once the baby leaves the hospital. Similarly, much expenditure for the aged occurs at the very end of life while too little is invested in enhancing functioning and independence of this population. The large uncertainty here is how we reallocate our current investments for improved health and welfare for children and the elderly, without attracting entire new populations demanding these services.

The process of making such decisions may be as important as the decisions themselves. Health providers and population groups are polarized in many ways, and as new knowledge and technology present us with profound ethical and economic choices on which reasonable people differ, we will need more elaborate frameworks and procedures for gathering information, hearing representative viewpoints, and achieving a resolution seen as fair and equitable. Components of such a process exist—such as special studies, technology assessments, disinterested panels, consensus conferences, and the like—but we have to devote much thought to developing mechanisms perceived as legitimate in arriving at such difficult decisions.

It is clear that new services and technologies, once introduced, are extraordinarily difficult to withdraw if they promise any advantage at all. While many technologies may have costs beyond any expected benefits, and may be applied in highly wasteful and inappropriate ways, the fact that they have some marginal advantage in particular

instances is used to justify their widespread diffusion. This would argue for better control over the diffusion of new procedures and technologies and improved processes for decision making on reimbursable components of care. The earlier such decision-making processes are initiated, the better the chances of successful resolution without thwarting or delaying the introduction of efficacious care.

While we face a future of many uncertainties, the existing ferment also offers new opportunities. Subspecialties, profit and nonprofit institutions, insurance programs, and varying professional groups are more vigorously competing for market shares than ever before. While their vigorous advocacy and sophisticated communication and organizational skills make them formidable actors in the public arena, the sense of a newly emerging format of care for the future make these groups more innovative and more willing to bargain. The professionals, voluntary hospitals, major insurance plans, and proprietary institutions and industries will carefully look after their own interests. The public interest, however, requires a system of alternatives where results are structured to provide access and care of comparable quality to our most needy and unfortunate groups, and particularly to those that lack the influence and sophistication to insure their future prospects. There can be no higher goal than to ensure that persons regardless of race, income, or origin have access to health services that provide decent opportunities for achievement of personal aspirations.

References

- Aaron, J.H., and W.B. Schwartz. 1984. *The Painful Prescription: Rationing Health Care*. Washington: Brookings Institution.
- Brook, R.H., J.E. Ware, Jr., W.H. Rogers, E.B. Keeler, A.R. Davies, C.A. Donald, G.A. Goldberg, K.N. Lohr, P.C. Masthay, and J.P. Newhouse. 1983. Does Free Care Improve Adults' Health?: Results from a Randomized Controlled Trial. *New England Journal of Medicine* 309:1426-34.
- Freshnock, L. 1984. *Physician and Public Opinion on Health Care Issues: 1984*. Chicago: Survey and Opinion Research Section, American Medical Association.
- Fried, C. 1975. Rights and Health Care: Beyond Equity and Efficiency. *New England Journal of Medicine* 293:241-45.
- Friedson, E. 1970. *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. New York: Dodd, Mead.

- . 1975. *Doctoring Together: A Study of Professional Social Control*. New York: Elsevier.
- Hamm, L., T. Kickham, and D. Cutler. 1982. Research Demonstrations and Evaluations. In *Long-term Care: Perspectives from Research and Demonstrations*, ed. R. Vogel and H. Palmer, 167–253. Washington: U.S. Department of Health and Human Services, Health Care Financing Administration.
- Iglehart, J. 1984a. Opinion Polls on Health Care. *New England Journal of Medicine* 310:1616–20.
- . 1984b. The Twin Cities' Medical Marketplace. *New England Journal of Medicine* 311:343–48.
- Klein, R. 1983. *Politics of the National Health Service*. New York: Longwood.
- Luft, H. 1981. *Health Maintenance Organizations: Dimensions of Performance*. New York: Wiley-Interscience.
- Lurie, N., N.B. Ward, M.F. Schapiro, and R.H. Brook. 1984. Termination from Medi-Cal: Does It Affect Health? *New England Journal of Medicine* 311:480–84.
- Manning, W., A. Leibowitz, G.A. Goldberg, W.H. Rogers, and J.P. Newhouse. 1984. A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services. *New England Journal of Medicine* 310:1505–10.
- Mechanic, D. 1976. *The Growth of Bureaucratic Medicine: An Inquiry into the Dynamics of Patient Behavior and the Organization of Medical Care*. New York: Wiley-Interscience.
- . 1978. Ethical Problems in the Delivery of Health Services. In *Appendix to Ethical Guidelines for the Delivery of Health Services by DHEW*. U.S. Department of Health and Human Services. DHEW pub. no. (05) 78-0011. Washington.
- . 1979. *Future Issues in Health Care: Social Policy and the Rationing of Medical Services*. New York: Free Press.
- . 1984. The Transformation of Health Providers. *Health Affairs* (Spring):65–72.
- Newhouse, J.P., W.G. Manning, C.N. Morris, L.L. Orr, N. Duan, E.B. Keeler, A. Leibowitz, K.H. Marquis, M.S. Marquis, C.E. Phelps, and R.H. Brook. 1981. Some Interim Results from a Controlled Trial of Cost-sharing in Health Insurance. *New England Journal of Medicine* 305:1501–7.
- On Lok Senior Health Services. 1983. *On Lok's CCODA: A Cost Competitive Model of Community-based Long-term Care*. San Francisco. (Mimeo.)
- Rice, D., and J. Feldman. 1983. Living Longer in the United States: Demographic Changes and Health Needs of the Elderly. *Milbank Memorial Fund Quarterly/Health and Society* 61(3):362–96.

- Sidel, V.W., and R. Sidel. 1983. *A Healthy State*. New York: Pantheon.
- Starr, P. 1982. *The Social Transformation of American Medicine*. New York: Basic Books.
- U.S. Department of Health and Human Services. 1984. *Health—United States 1983*. PHS pub. no. 84-1232. Washington.
- Wennberg, J., and A. Gittelsohn. 1982. Variations in Medical Care among Small Areas. *Scientific American* 246(4):120–33.
- Wennberg, J., K. McPherson, and P. Caper. 1984. Will Payment Based on Diagnosis-related Groups Control Hospital Costs? *New England Journal of Medicine* 311:295–300.

Acknowledgment: This work was supported, in part, by a grant from the Robert Wood Johnson Foundation.

Address Correspondence to: David Mechanic, Ph.D., Medical Sociology Research Unit, Room 203, Murray Hall, Rutgers University, New Brunswick, NJ 08903.