
Economic and Social Research Institute

Issues in Coverage Expansion Design

History and Primer on Cost Containment Efforts and Implications for Future Prospects

by Elliot K. Wicks, Ph. D.

Economic and Social Research Institute

1015 18th Street, N.W., Suite 210
Washington, D.C. 20036
(202) 833-8877
www.esresearch.org

April 2004
Number 6

Covering
America

REAL REMEDIES
FOR THE UNINSURED

Sooner or later, anyone seriously proposing major reforms to move the country toward universal health coverage has to confront the problem of health care cost escalation. Health care expenditures in the United States have continued to rise over the years, both in absolute terms and as a share of total goods and services produced in the country (see figures below). At no point between 1960 and 2001 was the annual rate of increase in health expenditures less than 5 percent. Recent resumption of the rapid cost escalation that was the rule except for the brief respite of the mid-to-late 1990s has made everyone aware of the urgent need to control health costs. Since any expansion of health coverage is certain to add to the total demand for medical services, those who propose to expand coverage need to address the issue of how to keep health care costs under control.

The purpose of this paper is to outline the range of cost control options from which policy reformers might choose. The United States has a long history of trying various costs containment strategies, none of which

has proved very successful. It is therefore useful to review what has been tried and to consider all the available options.

We begin by asking why controlling health costs is such a problem. After all, we do not talk about the need for public policies to control the costs of automobiles, computers, theater tickets, or, for that matter, most other goods and services we consume. We let market forces operate, and we generally find the results to be acceptable. Why is health care different? There are a number of characteristics of medical services that make them different from most other products and that help account for the cost control problem and the failure of market forces to constrain costs.

A key factor is consumers' limited understanding about their needs for medical care and their lack of knowledge about the specific services and products that best meet their needs. Economists say that markets work well only when information about the price and nature of the product is readily available to consumers so that

they can weigh the relative costs and benefits. That is seldom the case for health care. We seek the counsel of doctors and other health professionals precisely because we want them to fill in our knowledge gaps. They diagnose what is wrong and tell us what products and services we need to “consume” to address the problem, and we often know little if anything about the cost until after the care is provided.

With the recent availability of medical information on the Internet and elsewhere, some health care consumers may be more knowledgeable than in the past. But for most people, the highly technical nature of modern medicine, along with the rapid advances, make dependence on the providers’ advice mandatory.¹

But of course, doctors and other providers are the suppliers of medical services. Usually, the more goods or services suppliers provide, the higher their income. If the market for health services were allowed to operate as most markets do—and as health markets often have—suppliers could determine demand. This is essentially the situation that occurs when health care providers are paid on a fee-for-service basis. At the very least, the incentives under these circumstances do not encourage suppliers to economize.

There is a long discussion in the health economics literature about the nature of “provider-induced demand” and its cost-inflating effects. The gist of that literature is that when there is excess supply among providers, the demand tends to rise until excess capacity is substantially reduced—a result that would be unlikely if providers did not strongly influence the demand for medical services.

A second major cause of cost escalation in medical markets is the presence of health insurance. People whose medical expenses are covered by insurance pay far less than the full costs of the services they consume (at least at the time of service; they pay indirectly when they pay premiums). They have much less reason to be attentive to the cost or to weigh costs against benefits. In fact, once co-pays and deductibles are met, the price of covered services is zero. Because consumers pay less than the production cost of the goods they consume, too many resources get allocated to production of health services. And, of course, cost escalation pressures are exacerbated.

A third significant cause of health cost escalation is rapid technological change, accounting for somewhere between one-half to two-thirds of long-run cost trends.² Medicine continues to make major advances in capacities to diagnose, treat, and cure diseases and injuries, and usually patients benefit in the form of greater well-being. The rate of technological advance has, if anything, picked up in recent years, and no end is in sight. The new technologies are typically more costly than those they replace, and the improvements in health status often come at a high marginal cost. The problem is that our willingness to consume these new technologies is virtually limitless. We feel we should have access to any medical service that has the potential to appreciably improve our well-being.

Other factors contribute to cost pressures:

- lessening competition among some categories of providers, especially hospitals and certain physician specialties, which is a result of barriers to entry;

- the relatively slow pace with which best medical practices are disseminated and adopted; and
- aging of the population, although this is not a major factor.³

The upshot of all these peculiarities of health care is that free market forces have not produced very good performance with respect to controlling costs. Most students of health care financing agree that some degree of intervention is necessary to keep costs in line, though there is much disagreement about the form that intervention should take.

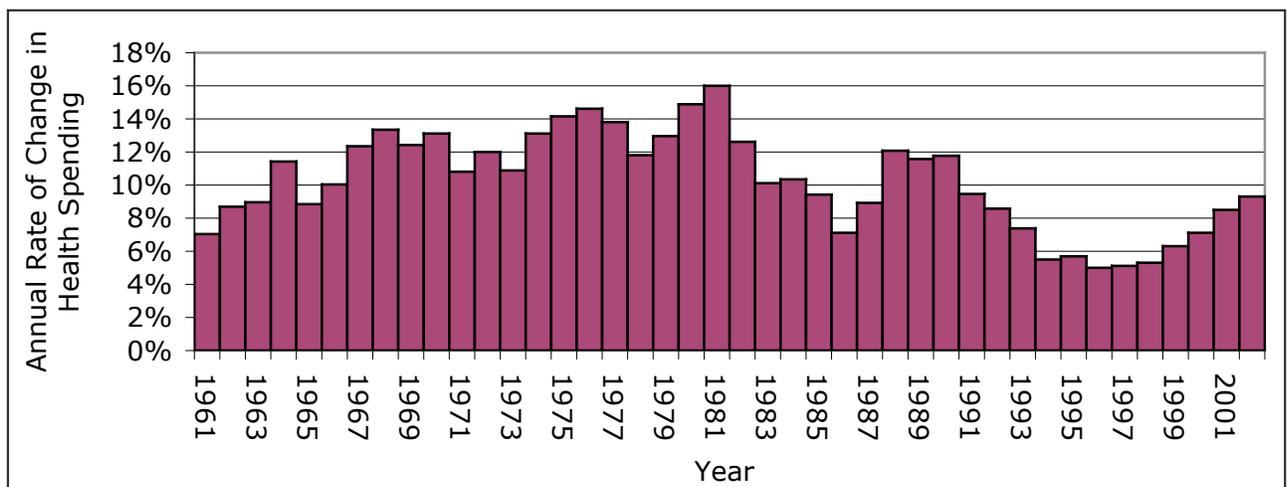
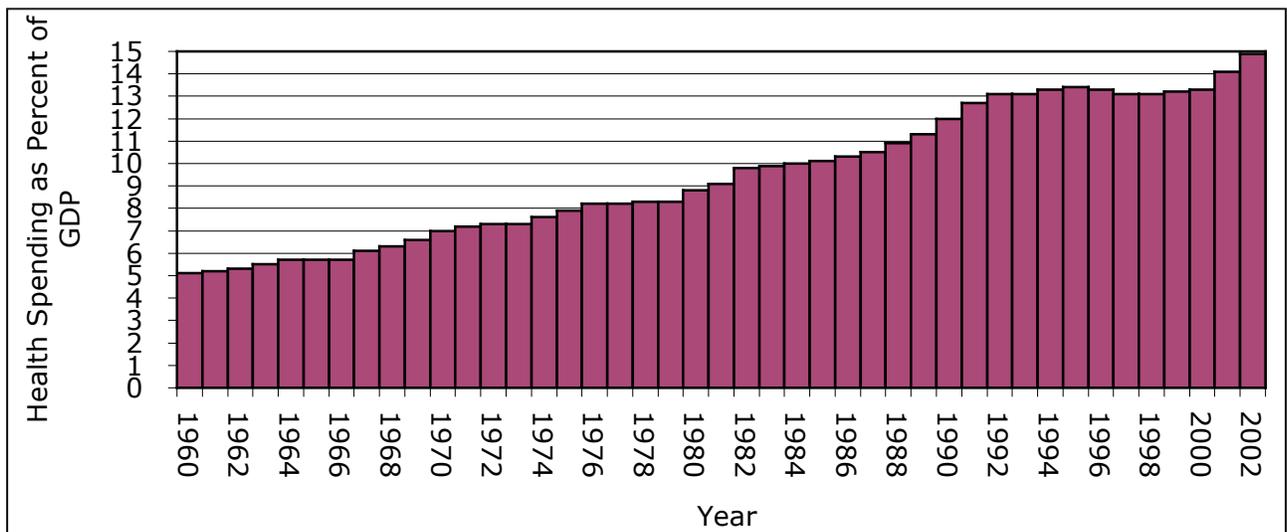
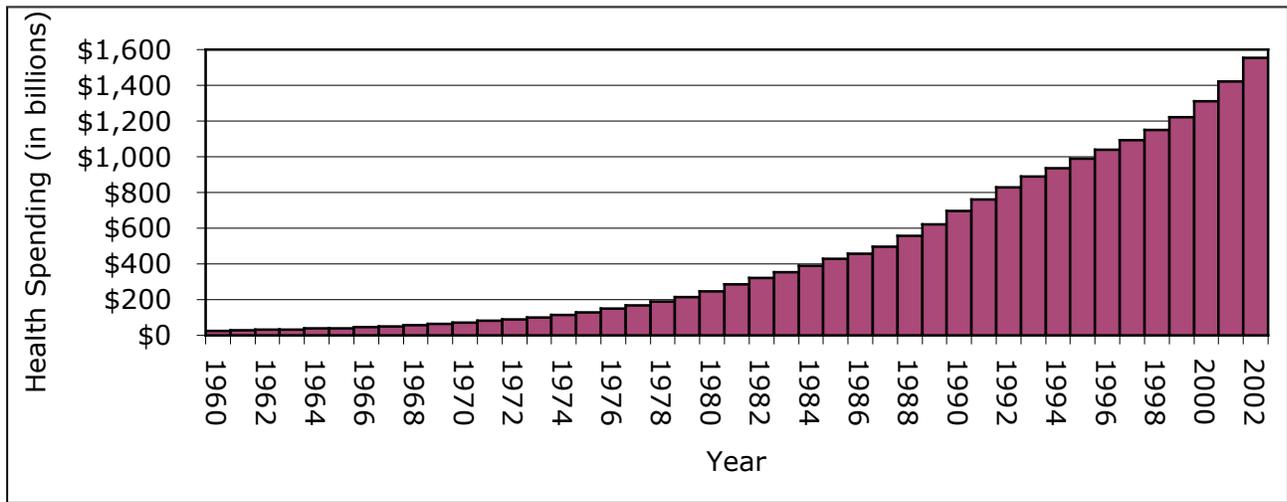
Before looking at specific strategies to contain costs, it is useful to look at the problem from a conceptual point of view. Conceptually, there are only three ways to control costs: produce and consume fewer services, pay providers of those services less per unit of production, or improve efficiency of production. None of these approaches is painless or easy to achieve. The first—consumption of fewer services—is likely to provoke the ire of consumers and providers and raise the specter of rationing, unless the constraint is achieved by cutting the consumption of medical services that are perceived as having almost no value, which are hard to find. The second—paying providers less—is obviously going to be opposed by providers, since their income is diminished, and there are limits on how far it is possible to go in this direction and still have a sufficient supply of providers willing to offer services. The third—improving efficiency of production—seems least likely to generate opposition, but even this approach yields lower income for someone. As Uwe Reinhardt has often observed, what are seen as costs by those who pay medical bills are income to somebody in the medical care system. No wonder cost containment is not easy.

¹ Ha T. Tu, J. Lee Hargraves “Seeking Health Care Information: Most Consumers Still on the Sidelines,” Issue Brief No. 61, Center for Studying Health System Change, March 2003

² Paul B. Ginsburg, presentation at “Health Care Costs and Health Coverage,” sponsored by the Alliance for Health Reform, Washington, D.C., Aug. 2, 2002.

³ In 2001, aging of the population accounted for only 8.7 percent of the increase in costs. Ginsburg, 2002.

Changes in Health Care Spending, 1960-2002



Source: Department of Health and Human Services, Centers for Medicare & Medicaid Services, Office of the Actuary: National Health Statistics Group; U.S. Department of Commerce, Bureau of Economic Analysis; and U.S. Bureau of the Census.

Debates about cost containment often used to be posed as a choice between competition or regulation—letting the market work or having government intervene in markets in some way. But this categorization is not very useful, for several reasons. First, since federal, state, and local governments are major payers, accounting for more than 45 percent of total expenditures, there is not any question that the medical system will be subject to some considerable government regulation. Secondly, the categorization of regulation vs. competition greatly over-simplifies the nature of options that are worthy of consideration and tends to make people reject some options on the basis of their categorization without giving them objective consideration.

We turn now to the range of cost containment strategies that are available.

Fee-for-service payment and unfettered markets

One approach, at least in theory, is to allow markets to work without constraints and to depend on provider competition and informed, cost-conscious consumers to keep costs under control. Under this approach, providers charge whatever the market will bear and are paid on a fee-for-service basis. This is the way most markets work, but the modern health care economy has never really operated exclusively on this basis. Because of the peculiar characteristics of health services and the distorting effects of insurance, described earlier, the modern health care economy has always operated with some constraints on provider payments. Though for many years, fee-for-service payment to doctors and other individual practitioners and cost-based reimbursement for hospitals was the rule, public and private insurers always placed some limits on

what they would pay, as outlined in the next section.

Administered prices

Constraints on Fee-for-Service

Third-party payers—public and private insurers—have almost always imposed some rules that limited how much they would pay providers, presumably because they understood that heavily insured consumers are not likely to be vigilant in avoiding excessive provider fees. Instead of letting market forces set prices, insurers established “administered price” systems, which limited the amount they would pay (though sometimes the patient had to pay the excess above the “fee schedule” amount). Various approaches were used in the early years to set these fees, one of the most common being some variant of the “usual, customary, and reasonable” (UCR) approach. Essentially, the UCR approach involved a formula that set each doctor’s fee by assigning weights to the amount that the doctor usually charged for a particular procedure (usual charge) and to the amount other doctors were charging for the same procedure (customary charge), and with some overall constraint on the rate of increase over time (reasonable charge). The idea was to allow some flexibility in fees among doctors, while keeping the differences within a reasonable range to ensure fairness, and to allow for some increases in fees over time but at the same time to prevent them from escalating too rapidly. Unfortunately, the approach did not work very well. Individual physicians and physicians as a group could influence their fees simply by continually charging more, which would gradually raise the amount they actually got paid. The system was not very fair, either, because physicians who increased their fees regularly received more income than those who showed

more restraint. Besides, the administered price approach affected only the price side of the price/quantity cost equation. It did nothing to discourage physician-induced demand, and it provided no incentive for physicians to economize on the use of other medical resources.

Cost-based reimbursement for hospitals had the same fault. As long as hospitals were paid for every service they provided at a rate sufficient to cover the costs, whatever they were, incentives to economize were completely absent.

The federal government led other third-party payers in recognizing that it had to come up with different hospital and physician payment systems if costs were to be controlled. Medicare decision-makers saw the need to build in incentives for providers to economize on the resources they used to provide health care services.

Physician Payment - RBRVS

Dissatisfaction with the UCR system for paying physicians led Medicare to develop a new, more objective way to set fee schedules. Instead of basing the payment rates on historical fees, Medicare undertook efforts to develop a fee schedule that would reflect the real resource costs incurred in producing services. The Resource-Based Relative Value Scale (RBRVS), initially implemented in 1992, was the result. In this system, a weight is assigned to a particular simple, base-line physician procedure, based on the associated costs, including physician work, practice expenses, and malpractice insurance costs. Other, more complicated procedures are then assigned a higher value as a multiple of the base-line procedure, depending on the additional work involved. Compared to UCR payment, this approach is more objectively based, attempting to emulate the fees that would naturally occur in a well-functioning, competitive market, where fees reflect real resource costs.

Inefficient physicians cannot make up for their inefficiency simply by charging higher fees; so there is a built-in incentive for physicians to use their own resources efficiently. Yet the approach still does nothing to discourage high utilization: it is still true that physician income increases in direct proportion to the number of services provided. And it does nothing to encourage physicians to economize in choosing the mix of other medical resources they order on behalf of patients

Hospital Rate Regulation

Early attempts to constrain hospital rate increases focused on regulating hospital prices. Paid on what was essentially a cost-plus basis, hospitals had no financial incentives to hold down costs. Moreover, in most markets there were not enough hospitals to produce effective price competition even if hospitals had been paid on something other than a cost-plus basis. In the eyes of some critics, this made hospitals like public utilities and led them to the conclusion that their prices should be regulated, as prices of public utilities were regulated. For these and other reasons, a number of states looked to hospital rate regulation as a solution. If rates were limited by regulation, the argument went, hospitals would be forced to keep their costs low enough to avoid incurring financial losses. Although there was a flurry of interest in this approach to cost control in the late 1970s, only a few states adopted this approach, and only Maryland continues to operate under a hospital rate setting system. One problem, besides strong opposition from hospitals, was that limiting price increases while still operating in a fee-for-service system did not create incentives to limit utilization rates. Hospitals could add revenue by providing more hospital days—more admissions and longer stays.

The Voluntary Effort

One interesting variant on this rate-limiting approach was the so-called “Voluntary Effort” of 1977. As part of a large effort to bring inflation under some modicum of control, the Carter Administration proposed to regulate hospital costs. The hospital industry agreed instead to voluntarily limit cost increases to specified target rates. This was price control but voluntary compliance. The target was met only for the first year.⁴ Without any enforcement mechanism, or at least the imminent threat of imposed controls if escalation continued, it was almost inevitable that hospitals would respond to continued cost pressures on them by the raising prices they charged their customers.

Expenditure Targets and Fee Schedules

Though never tried on a widespread basis in the United States, some European countries that used fee schedules, for example, Germany, experimented with establishing “targets” or caps on total expenditures for a provider group—for example, total physician expenditures—and then adjusting the fees downward if the targets were exceeded. Sometimes, they would also *increase* fees if expenditures came in below the target. While this approach penalized the provider group for “excess” expenses by the group as a whole, it really provided no incentives for individual providers to contain utilization. Individual physicians could maximize their income by providing many procedures. Even though such behavior would contribute to expenditure levels that exceeded the target, the indi-

vidual physician had no incentive to keep his or her utilization rates down—in fact, quite the contrary. The physician who behaved “well” and kept utilization down would be penalized, both by having lower income in the present and suffering later when the fee schedules were lowered.

Incentives for Cost Containment

Provider Incentives: Payment for “Bundled” Services

As people pondered this problem, it became clear that some provider payment approach was needed that would create financial incentives to economize on the use of services. Individual payment for every single separate service, no matter how the payment rate is determined, provides no incentive to choose the most efficient mix of services to treat a patient. A way to address this problem is to pay for larger “bundles” of services. For example, if hospitals were paid a fixed amount for a whole day of patient care (per diem payment) rather than for every service item, they would have incentives to economize on all the elements of service that go into providing a day of care. But they would still get rewarded for providing more days of care, that is, for higher admission rates and longer lengths of stay. The solution appears to be to pay for the largest bundle of services possible. For example, instead of paying a hospital on a per diem basis, payment might be based on an entire hospital stay, which provides incentives to reduce length of stay (though not the number of admissions). At one level higher, hospitals might be paid for an entire episode of care, including inpatient and outpatient services. Or at the highest level, a provider (for example, an HMO) might be paid for providing all of the care needed for a defined num-

⁴ Karen Davis, “Recent Trends in Hospital Costs: Failure of the Voluntary Effort,” Testimony before the U.S. House of Representatives, Committee on Energy and Commerce, Subcommittee on Health and the Environment, Hearings on “Increase in Hospital Costs: Is the Voluntary Effort Working,” Dec. 15, 1981.

ber of enrollees. This approach is, of course, capitation payment.

The practical problems of implementing the concept of paying for bundled services are great. Perhaps the most difficult problem is how to determine a rate for the bundle of services that is enough but not too much—that is, an amount just sufficient to cover the full costs of providing the needed services in an efficient way. If the payment is too little, even efficient providers will not be able to cover costs. If the payment is too much, the incentives to be efficient are not as strong and health costs are higher than necessary. Determining the correct payment rate is especially difficult because the cost of providing a bundle of services, however defined, can vary greatly with the severity of the patient's condition. The resources required to provide a day of hospital care in intensive care, for instance, are much greater than those associated with caring for a mother in her last day of a routine maternity stay. The same problem of accounting for differences in patient resource needs arises no matter how the bundle of services is defined. Finding ways to accurately adjust for case-mix or illness severity is a great challenge, and no matter what the methodology, its application is almost always controversial. Keeping the payment amounts at the right level as technologies change—which changes the optimum mix of resources—is also very challenging. Additional problems arise because the cost of resources differs from region to region. Making sure that providers do not categorize patients into a higher severity category to increase payment levels (“upcoding”) is a further challenge.

The first widespread use of payment for bundled services was Medicare's adoption in 1984 of the DRG (diagnosis related group) payment for hospitals. The concept involves deciding how much, on average, hospi-

tals spend to treat patients admitted for each of a large number of diagnoses. That amount then becomes the amount that each hospital receives for caring for a patient with that diagnosis (the DRG payment). Hospitals able to provide the needed care for less than the set (average) amount are rewarded with profits or an addition to surplus. Hospitals whose costs exceed the allowed amount either have to change their ways or sustain long-term losses. So all hospitals have incentives to find the most economical mix of resources to treat the patient during his or her stay. This system provides incentives to reduce length of stay and economize on resources during the stay, but it does not include incentives to reduce the number of admissions. But the experience with the DRG system as it was introduced was that admissions also fell. The system continues in force, though with many refinements, as Medicare's payment system for hospital care.

The other prominent use of services bundles is capitation payment for HMOs. In the archetypical case, the HMO is the provider and is obligated to provide all needed services for a defined population in exchange for a fixed payment for each person enrolled with the HMO. The HMO as provider is at risk for the whole range of care. Faced with this cap on total income, the HMO has incentives to use the most economical mix of resources, to constrain payments to providers, and to discourage excess utilization of services. In fact, it was the early research that showed that the capitated prepaid group-practice plans like Kaiser Permanente and Group Health of Puget Sound had much lower utilization rates for hospital care and lower total costs that lead the Nixon administration to support successful legislation to promote the growth of capitated health plans. Indeed, it was this 1973 law that introduced the term “health

maintenance organization” into the health care lexicon.

Of course, as managed care plans have evolved away from the original group practice plans so that the insuring organization is often not the direct provider (as with independent provider association HMOs, or IPAs), the distinction between capitation and an ordinary insurance premium paid to an insurer is blurred. In either case, the insurer is obligated to pay for necessary services for a period for a fixed payment per enrollee, but in the latter case, the insurer's proximity to, and ability to control the use of, the medical services is more tenuous. Of course, it is common for managed care plans to pass on the risk by capitating sets of providers, who then have to provide all necessary services falling within a defined bundle of services (for example, ambulatory care) for a fixed payment per person signed up with the provider.

At least in some respects, the experience with bundling services through capitation did prove successful. Most observers attribute the period of relatively slow health care cost growth of the mid to late 1990s to the rapid growth of managed care plans (although this period of cost stability was probably also a result of some “underpricing” of premiums, reflecting insurers' attempts to “buy” market share during a time of intense competition). The growth of managed care caused utilization rates for inpatient hospital care and some procedures to decline. But it also produced a backlash, as consumers rebelled against what they saw as denials of access to care they had come to expect. In any case, most students of health care financing agree that it was easier for managed care plans to lower the expenditure base than to reign in the rate of growth of expenditures once the initial savings was realized.

Provider Incentives—Paying for Cost-Effective Care

In theory, at least, it would be possible to link physician or hospital payments to some output measure that takes account of both cost and quality. Providers who can demonstrate they can provide cost-effective care would get more than providers offering a less ideal combination of high-quality and lower-cost services. Real world examples are hard to find, in part because of the great difficulty of measuring quality and the previously discussed problem of accounting for differences in patient severity.

Consumer Incentives

Although the peculiar characteristics of the health care market have caused those concerned with cost escalation to focus on developing incentives for providers, many people believe that consumer incentives can also have an economizing effect.

Increased Consumer Cost Sharing

As noted earlier, insurance coverage reduces the price of medical services and encourages consumers to buy more than they would if they had to pay the full cost out of pocket. Over the years, as health insurance has become the predominant method of purchasing medical care, coverage has become more comprehensive; it covers more services and pays more of the total bill. But the less people pay, the weaker the incentives to economize. An obvious solution seems to be to change the system so that consumers pay more of the price.

Greater Consumer Cost Sharing—Selective Co-Payments

In just the last few years, health plans and employers have experimented with a form of cost sharing that gives consumers incentives to choose less expensive medical options when substitutes are available. The

most widely used form is the “tiered” prescription drug co-payment approach. The co-payment is lowest for the generic version of a drug. When a patient uses a brand name equivalent that is on the insurer’s formulary, the co-payment is higher. And when the drug is a brand name not on the formulary, the co-payment is even higher.

Although this tiered co-payment approach has to this point been used primarily for prescription drugs, there is interest in applying the concept to other medical purchases that consumers make, such as the choice of a hospital or medical group. If an insurer has identified certain providers or facilities as being particularly cost-effective, they could then assess lower co-payments when patients choose to utilize those providers or facilities. To make this approach defensible, however, insurers have to show that they are using measures other than just cost to justify favoring some providers over others.

Greater Consumer Cost Sharing—High-Deductible Plans

Another cost sharing approach is to impose higher deductibles—an amount that consumers must spend out of pocket, frequently for the whole range of covered services, before the insurance kicks in. For example, a policy might require each individual family member to spend \$150 or the family to spend \$500 (whichever occurs first) before the insurer will pay anything. Since the consumer is paying for everything until the deductible is met, he or she has a strong incentive to economize.

Many people have argued that the logical way to approach this problem is to alter the health insurance system to make deductibles much larger, in the thousands of dollars rather than the hundreds. They point out that the purpose of *insurance* is to protect people against *large, unpredictable* losses, not to prepay for small, predictable

expenses that are budgetable. Homeowners’ insurance pays when one’s house burns down; it does not pay to repair or replace a malfunctioning dishwasher. If it did, people would replace dishwashers much more frequently. If deductibles were large—if health coverage were limited to paying for medical catastrophes—most people would be paying for most of their care out of pocket, since most people do not consume large amounts of medical care in any year. And that would give them strong incentives to care about the costs of care.

Critics of this approach raise a number of objections:

- High deductibles and catastrophic coverage would likely have limited effects on the most expensive medical decisions. The rough rule of thumb is that in any population, about 70 percent of the expenses in a year are accounted for by about 10 percent of the population. Most of this 10 percent are very ill people who would have used up any deductible; and after that, if they have good catastrophic coverage with high lifetime maximums, they have no financial incentive to worry about cost. So the cost-saving potential of heavy cost sharing is limited, since it does not much affect decisions for the most expensive episodes of care.
- This high cost-sharing approach is inconsistent with the evidence that physicians strongly influence most expensive medical decisions. These decisions are made at times of medical crisis, when people are not psychologically prepared to carefully weigh costs against health benefits and are most likely to rely on the advice of a trusted physician.
- High consumer cost sharing might discourage people from getting needed care, especially preventive care. In an unusual social experiment conducted in six cities from 1974 to 1982, the Rand Health Insurance Experiment randomly assigned people to health plans with different levels of

cost sharing. The study found that cost sharing did have a significant effect in reducing use of medical services but that it was as likely to reduce the use of needed services as unneeded services, and that it was particularly likely to reduce utilization by people with certain kinds of medical problems.

- Most people who buy coverage do not want catastrophic benefit plans. They want comprehensive coverage. They want to be covered for ordinary, more-or-less routine expenses, not just the big-ticket items they are unlikely to need. Efforts to sell catastrophic coverage have generally not found many buyers. If we were starting from a point where no one had had comprehensive policies with first-dollar coverage, catastrophic plans might be better sellers. But once people get used to having coverage that pays for a high proportion of most services, it is hard to convince them that a high-deductible plan is good coverage.

- The people most likely to buy high-deductible coverage are lower-risk, higher-income people who want to protect themselves against loss of assets, but who do not anticipate needing much care. The people most likely to prefer comprehensive, low-deductible plans are people with modest incomes, for whom any significant expenditures may be a burden, and people who anticipate needing substantial amounts of care, especially those with chronic conditions. The consequence could be that the low-risk people congregate in one insurance pool, while the high-risk people are in another. Without joint pooling of high- and low-risk people, the disparities in premiums are likely to grow very wide, making coverage less affordable for lower-income people.

Medical Savings Accounts

Sensitive to these criticisms, some of the people who favor the incentives

inherent in catastrophic or high-deductible benefit plans have devised a modification known as medical savings accounts (MSAs). The same basic concept is inherent in the newer “consumer-driven health plan” movement. In simplest form, the idea is to combine catastrophic coverage with an account, to which the employer contributes, that could be used to pay the first-dollar medical costs before the high deductible is met. This approach is designed to address several of the criticisms above. People covered by such a plan do not have to come up with cash entirely out of their normal financial resources to pay for routine care; they use the money in the medical savings account instead. Since the money is already there, they may also be less likely to defer or avoid consuming preventive services. But since any money they do not use is theirs (not the insurer’s)—perhaps ultimately to be used for non-medical expenses—consumers have incentives to economize. In addition, since there is money to pay for first-dollar services, people may be more enthusiastic about this option than they are about stand-alone catastrophic coverage.

Although MSAs and consumer-drive plans address some of the criticisms of the high-cost sharing approach to consumer incentives, the other criticism remain to be addressed. Moreover, the experience to this point suggests that consumers are not enthusiastic about this kind of coverage. MSAs represent only a tiny portion of all benefit plan types sold.

Managed Competition

High-deductible plans give consumers incentives to economize at the point where they consume medical services. A different approach to encouraging consumer cost consciousness is to give people incentives to be economical at the point where they choose a health plan. People who espouse this approach, perhaps most

prominently Alain Enthoven and earlier Walter McClure, acknowledge all the previously discussed difficulties involved in trying to get consumers to make careful cost-benefit calculations when they actually consume medical services. They say that it is an ineffective approach. Instead, they propose to set up an incentive structure that would motivate consumers to compare health plans and select high-value plans, which by creating competition among health plans, would force them to become more efficient. The notion is that health plans, and their affiliated providers, have the expertise to deliver efficient care if they have sufficient incentives to do so.

This theory behind this approach requires certain conditions:

- Consumers must have a choice among a number of health plans.
- Consumers need to have strong incentives to choose high-value plans. Ideally, in setting premium contribution policies, employers would contribute a fixed-dollar amount (not more than the cost of the least costly plan), regardless of the health plan the employee chooses. Thus, in deciding to buy a plan that is more expensive than the least costly plan, employees would be spending entirely their own money and would select such a plan only if they were convinced that the extra value justified the extra cost.
- Health plans would differ in significant ways—with respect to their networks of providers, their “style” of medical practice, the degree to which they restrict access to services, their emphasis on quality and good outcomes, etc.
- Consumers would have a source of reliable information and the means to compare health plan performance.

Managed competition has not had outstanding success as a cost containment tool, at least not yet. In the eyes of proponents, one reason is that it has not really been tried. Too many

employers still offer no choice of plans, and when they do, they often pay the entire premium, or an amount that is substantially more than the cost of the least costly plan, or the same percentage of the premium regardless of the plan cost. When employers have no plan choice and/or weak financial incentives, they are unlikely to select the most cost-effective health plans; so what competition there is among plans is not a result of vying for individual consumers' business based on price-value comparisons.

The theory of managed competition was also based on the notion that managed care plans would try to distinguish themselves by offering higher-value care. They would be selective in choosing their provider panels and include only providers with a track record, or at least a strong prospect, of providing high-quality, cost-effective care. But that did not happen. As managed care plans began to compete, they found that the single thing that most concerned consumers in selecting one plan over another was being able to stay with "their" doctor, "their" hospital, or other provider with whom they had an established relationship. What employees care about, employers tend to care about. Thus employers were more prone to select health plans that had very large networks, so that they would not have to contend with employees complaining about having to change providers. The consequence in most areas of the country was that most health plans sought to sign up almost all providers, which meant that there was little difference among health plans in the most important characteristics—provider quality and the style of medical practice.

Of course, managed care plans might have evolved differently if employees had had meaningful choices among plans and strong financial incentives to choose less costly, high-

value plans. But even with those elements in place, consumers' strong loyalty to providers that have served them in the past might have made it difficult for plans to be narrowly selective in establishing their provider networks.

Finally, the buyers of health coverage, most notably employers, did not seem prepared to support managed care plans when they really sought to contain costs in ways that required consumers to change their expectations. Well-insured Americans have generally had virtually unlimited access to any care they thought they should have. As noted earlier, the options for containing cost are lower provider payments, provision of less care, and greater efficiency. Once managed care plans had largely exhausted the potential for provider discounts and turned to reducing utilization, they got much resistance from consumers and their doctors. People did not like having restricted access, having to wait to get approval for particular services, or being told that the some specific care was unnecessary. Employees complained to their employers, and instead of standing firm and educating their employees to understand that unlimited access and cost control are incompatible objectives, employers gave in. Health plans responding by doing what employers wanted, and pulled back on utilization controls. And politicians got involved. Under the banner of patient protection, they supported some measures that make even entirely defensible efforts to control costs difficult or even impossible to implement. Examples include "any willing providers laws," which make it impossible for health plans to establish selective networks, and restrictions on the way health plans could financially reward or penalize providers based on their performance in limiting costs.

Now, of course, managed care plans made some mistakes as they

tried to constrain access and control utilization. They sometimes used techniques that did not really save costs, and in some instance, the denials of care were not consistent with good quality. Perhaps only a minority really *managed* care effectively. But real cost containment is going to require some limits on access and use; it will not be painless. Unless employers (and politicians) are willing to defend managed care plans that employ effective techniques to limit cost, they cannot contain costs.

Whether better implementation of managed competition concepts can effectively control costs remains to be seen. But the barriers should not be underestimated.

Supply Constraints

Based on some rigorous research as well as much everyday observation, critics of the health system in the early 1970s became convinced that supply "creates" demand. They concluded that when there are "too many" hospital beds, "too many" physicians, and "too much" technology, these resources get used even when that use does little or nothing to improve health status. Hospitals need to fill beds and to use their technologies, and physicians need patients. Of course, this kind of "excess" supply could not exist for long if it were not for the peculiarities of the health economy outlined earlier that give suppliers considerable influence over demand.

To the critics who saw the system this way, it seemed logical to try to control costs by limiting supply. Since normal market forces were not producing a rational allocation of supply, the only real alternative appeared to be to make those decisions by analyzing the needs for services and then finding ways to make the supply match the needs. Limiting supply through planning and government regulation seemed particularly ap-

appropriate since government heavily subsidized training of providers in medical schools and funded much of the acquisition of medical facilities and equipment, as well as paying for much of the medical care produced with these resources.

Federally Mandated Health Planning

There were numerous attempts to estimate the need for physicians, particularly various kinds of specialists, who were thought to be too numerous relative to primary care physicians and a special source of excess utilization. Analysts showed the certain delivery systems, especially group-practice prepaid plans (later known as HMOs), were able to deliver quality care with far fewer hospitals days and fewer physicians. At a time when health care cost escalation seemed out of control, such analysis probably played a part in persuading Congress to pass legislation establishing a nation-wide system of "health planning." The law went into effect in 1975 and required each state, as well as regions within each state, to put in place health planning agencies whose job it was to analyze the health needs of their populations and then to develop a health resource plan for meeting those needs. The state health plan, which was based in part on the local health plans, had to be reviewed by the state legislature. A key strategy in this process was a mechanism to limit the supply of hospital facilities and technologies. States were required to develop a process for reviewing all hospital proposals for substantial facility expansion or acquisition of new equipment and to grant "certificates of need" only to those proposed projects deemed necessary. Projects that did not make the grade could not be built. One state, Michigan, went farther and passed legislation that put in place a mechanism for deciding how to close excess hospital beds.

The experience with health planning was not especially encouraging. Much effort went into the process in many states, and the deliberations around the health plan and the certificate of need activities were often highly visible and contentious. But researchers generally concluded that the process did not have much measurable effect on the rate at which capital expenditures took place. The evidence for savings was weak. In part, the lack of success can be blamed on the way health planning was structured. Although, by law, consumer representatives were in the majority on the health planning boards at the local and state levels, it was the provider representatives who attended meetings with greater regularity and who were better armed with arguments and data to support further capital expansion. The providers had the resources to pay professionals to assume this role, while consumers were less likely to have the training, expertise, and resources to be an effective countervailing force. In addition, since there was no requirement of an overall budget cap within which expenditure decisions had to be made, no trade-offs were required: approving one project did not preclude approving others. Under those circumstances, a case could nearly always be made that a project would serve at least some useful purpose. In addition, hospitals are major sources of jobs, and hospital expansion means more jobs. Hospitals are prominent community institutions to which many local people have loyalty, and communities often resist what they perceive as a threat to "their" institution. Thus communities were not always supportive of decisions to limit hospital expansions, renovations, or acquisition of new equipment even when approval seemed certain to increase costs and was unjustified in light of the availability of duplicative services nearby.

After a number of years, the national health planning legislation expired. Most states either eliminated or greatly scaled back their certificate of need process, though a few continue to operate review and approval processes today. Remnants of other aspects of health planning, particularly collection of hospital utilization and cost data, remain in some states.

Technology Assessment

An important cause of cost escalation is the introduction of new, expensive technologies. There is every reason to believe that technological advance will continue and that many of the advances will add to costs. Part of the problem is that technologies that were developed for one purpose often get used for many others. For example, many of the diagnostic scanning technologies like CAT scans and MRIs were designed for diagnosis of life-threatening conditions like brain tumors, where their diagnostic superiority might be critical to successful treatment. But once available, the technologies get used for many other purposes where their use may be less critical. They may provide benefits in terms of better health outcomes than less advanced and less expensive technologies, but only marginally better. There needs to be some way to decide whether the benefits of new technologies justify the cost.

The traditional technique for doing such assessment is cost-effectiveness analysis. A derivative of cost-benefit analysis, the basic idea is develop a ratio of benefits to costs but with the difference that the benefits are stated in terms of some measure of health outcomes rather than dollars, since it is difficult and controversial to assign dollar values to health outcomes. Cost-effectiveness analysis is particularly appropriate when comparing technologies that are designed to accomplish the same objectives. The analysis leads to a statement of what

outcomes are achieved per dollar spent.

Though theoretically rational, the approach involves many difficulties. One is that both the costs and the benefits change once the technologies are introduced. Greater use may produce economies of scale, and new uses may be developed for the technology. Assessing benefits in a scientific, measurable way is often difficult, expensive, and time-consuming. Technologies evolve so quickly that the research may be quickly outdated. Perhaps most problematic, making a judgment about the relative worth of technologies that are designed to produce *different* outcomes becomes very difficult. Without an overall budget cap within which expenditure decisions have to be kept, it becomes very hard to justify ruling out a technology that produces significant benefits even if the cost is very high.

Perhaps for these reasons, most technology assessment is confined to determining whether an emerging technology is safe and effective. This is not the same thing as deciding whether the benefits justify the cost.

Expenditure Budgets

Because of the difficulty of controlling costs by moving from bottom to top—that is, by trying to control all the individual elements as a way of controlling the total—for many years some people have advocated establishing expenditure budgets. The idea is to decide first what we want to spend and can afford for health care and set that as the expenditure budget and then find ways to ensure that the component costs do not add up to more than the total. Without a budget, it is hard to make decisions about individual expenditure elements because there are no trade-offs to be made. Spending on one use has no effect on the amount available for other uses. Virtually every business and every household makes expenditure decisions within an implicit or

explicit budget limit, constrained by the amount of revenue available.

Some people believe the only way to limit health care costs is to apply this same principle to medical spending.

Global Budgets

Applying this idea to health spending would require setting a spending limit at some high political level—by individual state governments or the federal government, for example—and then finding ways to make sure that the components of health spending are kept within that limit, commonly referred to as a “global budget.” Achieving that goal would be difficult enough to do even if all spending were ultimately paid for by a single payer (as with Medicare), but it is even more difficult in a system where there are many payers and no central agency that has the authority to limit individual expenditure decisions. The present health care system clearly falls into the later category. The difficulties of applying the global budget approach to such a system is one reason why some critics of that system support a “single payer” approach to providing health coverage. The global budget approach to health care cost containment would be easier to implement under such a system.

The problem with global budgets is that they are only a starting point for cost containment. They at least make explicit trade-offs necessary, but establishing such a budget does not by itself build the mechanisms for making those trade-offs. How to allocate resources—how to decide which possible uses for the available money will be funded—is still a very difficult problem, especially since it is a virtual certainty that this country will never have a public health system like that of Great Britain, where decisions of this sort are made by a government process. We will almost certainly continue to have a system that involves many different private payers,

providers, and consumers making expenditures decisions. Finding a way to ensure that all those individual spending decisions do not sum to more than the budget cap seems like an almost insurmountable task. At the very least, many other cost containment mechanisms would have to be put in place throughout the system.

The Oregon Plan

The most recent (and perhaps the most ambitious) experiment with global budgeting took place in Oregon in the early 1990s. Reformers in that state decided to combine the global budgeting approach with cost-effectiveness analysis to achieve rational resource allocation. They went through a detailed, elaborate, and widely publicized process to assess the cost-effectiveness ratios of medical procedures.

Coverage was to be provided for Medicaid recipients for all conditions above a threshold on the prioritized list, with the threshold being set each session by the state legislature on the basis of actuarial estimates and budgetary constraints. The threshold was also to serve as the minimum benefit package for private-sector insurance pools. To arrive at a preliminary version of the prioritized list of services, the Commission adopted a cost-benefit methodology that ranked conditions and treatments according to four factors: their cost, the net duration of benefit, physician estimates of the likelihood that treatment could alleviate symptoms or prevent death, and citizen views on the seriousness of symptoms and functional limitations. More than fifty physician panels met to develop estimates. Citizen values were obtained through both a telephone poll of 1,000 state residents and a series of community meetings and hearings organized by Oregon Health Decisions, a community service organization dedicated to obtain-

ing citizen input on ethical issues in health care.⁵

The preliminary ranking of conditions and treatments was revised several times, in part to correct what seemed like counter-intuitive rankings and in part to respond to federal Medicaid requirements. The product of all this work was a rank order list of over 700 “condition-treatment pairs,” with the highest valued services at the top of the list. The idea was that knowing how much there was to spend and the estimated need/demand for each service, it would be possible to draw a mark at the point in the list where the budget limit would be reached. Services farther down that point in the list would not be covered services for a particular budget period.

Originally, the reformers saw this system being used to guide both public and private insurers. But the approach has been used to this point only to implement budget limits for the Medicaid program.

Improving Quality and Effectiveness of Medical Practice

For many years, prominent researchers who have studied the delivery of care in our major health institutions have reached the same conclusion: much care that is provided to patients every day is wasteful, inappropriate, and harmful; and much beneficial, cost-effective care is not provided at all or not when needed. As one widely known physician who studies health care quality has observed, “We have . . . significant medication errors in 7 out of every 100 inpatients, tenfold or more variation in population-based rates of im-

portant surgical procedures, 30 percent overuse of advanced antibiotics, . . . [and] 50 percent or more underuse of effective and inexpensive medications for heart attacks and immunization for the elderly.”⁶

Poor quality care is expensive as well as being harmful. When people get sicker because of medical errors or because they fail to get needed care, costs rise. Of course, not all improvements to quality of care would lower costs—there is substantial underuse as well as overuse—but the net effect would likely be a reduction in costs, to say nothing of improved quality of life for patients.

Much work is being done to find effective ways to improve quality and effectiveness of care delivery. Some of this involves research to identify best practices and to develop practice guidelines to move forward toward evidence-based medicine. Another part involves getting the information out to physicians in an effective way. And still another part involves finding ways to induce physicians to change patterns of medical practice—to use the information that is available. These topics are major issues in themselves and beyond the scope of this paper. But any successful effort to contain costs must include mechanisms and incentives to improve quality of care and the practice of medicine. And the need to focus in on this problem will only grow more acute as the rate of technological change accelerates.

Health Behaviors

Finally, any discussion of cost containment would be incomplete without at least mentioning the widely accepted proposition that the most effective way to lower health care costs would be by getting people

to adopt more healthful lifestyles. Many of the maladies from which we suffer and for which we seek medical attention are either caused by or greatly exacerbated by the bad choices we make regarding health-related behaviors—smoking, substance use and abuse, diet, exercise, exposure to stress, amount of sleep, and so forth. The research clearly shows that people who choose healthful life styles have lower morbidity and live well longer. This is certainly reflected in lower health care costs. The challenge, of course, is to find effective ways to induce people to change their health behaviors.

Conclusion

It is the oldest of clichés to say that there is no single strategy that we can adopt that will “solve” the health care cost problem. In fact, the experience from previous efforts to control costs is generally discouraging. Some things have worked, but nothing has worked extremely well. In part, this is because we do not really want them to work. Ultimately, any long-range cost containment strategy will require *rationing* of care. Some people will not get all the care they want and perhaps not even all the care they need (in the sense that the care would produce real benefits). We find that hard to accept. Americans—and probably most of the rest of the people of the world—would be willing to consume health care in virtually unlimited amounts. Our willingness to take advantage of new technologies that extend life, enhance functioning, and improve quality of life knows no bounds, and the technologies that can do that are on the horizon. We have, for example, just scratched the surface in genetic research that holds out the promise for greatly improved health and well-being.

But our capacity to produce the medical services that allow us to benefit from new technologies is not

⁵ Richard Conviser, “A Brief History of the Oregon Health Plan and Its Features,” Office for Oregon Health Policy and Research, http://www.ohppr.state.or.us/out%20of%20folder/reading/index_reading.htm

⁶ David M. Berwick, MD, MPP, “Escape Fire: Lessons for the Future of Health Care,” Commonwealth Fund, 2002, pp. 19-20.

unlimited because resources used for medical care are not available for other desired uses. Trade-offs are inevitable, even in a growing economy. If we spend more for health care, we have less for education, defense, environmental improvement, consumer goods, and everything else. Americans have been willing to tolerate ever increasing growth in health expenditures as a proportion of total GDP, but at some point—perhaps when health spending approaches 25 percent—we will decide that is enough.

We can do much to reduce health care costs that does not require painful rationing of care—especially by becoming more efficient in care delivery and improving quality. But ultimately we have to accept the need for rationing, just as we do in every other

area of the economy. That is, we have to develop processes or mechanisms for deciding which medical services to provide under given circumstances and which to not provide under particular circumstances. For no other kind of good or service do we hold the view that people should be ensured of having access to ever-greater quantities of the good or service so long as the consumer receives *any* positive benefit. We cannot do that for health care either. We have to be willing to say that some things are not worth doing (and paying for through insurance) because the cost-benefit ratio is too high. The backlash against HMOs' effort to limit utilization suggests that this will not be an easy task.

Thus those who seek to control health care costs face two difficult challenges—finding effective tools to

limit cost growth and convincing the American public that constraints on their use of services are necessary. ■

The Covering America project promotes serious consideration of a diverse range of comprehensive proposals to provide affordable health coverage for millions of uninsured Americans. The project is coordinated by the Economic and Social Research Institute, a nonprofit, nonpartisan institute in Washington, D.C., and is made possible by a grant from the Robert Wood Johnson Foundation of Princeton, New Jersey. The Foundation does not endorse the findings of this or any independent research or policy project.