

Supply-Side and Demand-Side Cost Sharing in Health Care

Randall P. Ellis and Thomas G. McGuire

The central institutional feature in health markets is that the price paid by insured consumers when health care services are demanded can be set separately from the price paid to providers when services are supplied. This fact suggests two alternate strategies for controlling the costs of health care: demand-side cost sharing, where patients must pay more in co-payments or deductibles, and supply-side cost sharing, which seeks to alter the incentives of health care workers to provide certain services.

In broad terms, any health care financing system has three goals: protect consumers against the financial risk of health expenditures; promote efficient levels and types of health care services; and to be fair to consumers and providers (however fairness is defined). We review the rationale, limits, and comparative advantage of demand- and supply-side cost sharing in health care while primarily focusing on the short-run pursuit of the first two goals. We then turn briefly to the long-run issue of technology adoption, as well as the how supply- and demand-side cost sharing may affect the fairness of the health system.

When this paper was completed in July 1993, specific details of the proposed Clinton health care reforms had not yet been released. General features that have been leaked suggest that the reform proposal will include a mandated minimum benefit package; elimination of insurance exclusions based on preexisting conditions; a reliance upon competition between health plans (“managed competition”) to contain costs; mandatory employer participation in the health insurance system; and guaranteed access to health insurance pools

■ *Randall P. Ellis is Associate Professor of Economics and Thomas G. McGuire is Professor of Economics, both at Boston University, Boston, Massachusetts.*

for the unemployed and other individuals that are generally not able to get group insurance rates otherwise. Also rumored to be under consideration are the possibilities of eliminating the full tax exemption for employer and employee contributions to the premiums; shifting of some of the health premium burden from employers onto consumers; and restricting fees paid to providers, either through price controls, or “global budgeting.” With the exception of the last proposed reform, all of these features directly affect only the demand side of the market, changing the prices that consumers pay for health services or health insurance. None of these reforms directly relies upon changing supply-side cost sharing, the form of payment to providers. We argue below that supply-side cost sharing is a neglected component of national health care reform that may have an important role in cost containment and efficiency enhancement.

We intend our discussion to be at a very general level, and apply to demand- and supply-side payments to all types of providers. Because they represent the most ready application of the ideas we develop here, most of our examples focus on payments to hospitals. The roughly 5,000 institutions providing inpatient care are themselves a major part of the national health care picture, accounting for nearly 40 percent of national health care costs: an estimated \$359 billion out of \$903 billion in 1993 (Burner et al., 1992).

The Limits of Demand-Side Cost Sharing

Research in health economics in the 1970s was primarily concerned with how to best use one policy instrument—insurance coverage—to balance the achievement of the two goals of risk spreading and appropriate incentives to consume. To put it another way, insurance creates incentives for overconsumption of health care services, while limitations on insurance, whether in the form of deductibles, copayments, or coverage limits, force the consumer/patient generally to bear a greater risk of illness and monetary loss. This description of the literature also clarifies the fundamental (and unavoidable) limitation of demand-side cost sharing: since it is one instrument aiming at two conflicting goals, it can only achieve a second-best allocation.

Economists borrowed the insurance term “moral hazard” to describe the increased risk (or hazard) the insurer faces because of the presence of insurance itself (Arrow, 1963).¹ If health insurance is sold on the basis of how much medical care is utilized by uninsured patients, the insurer will be in for a nasty shock, since the existence of health insurance will lead to a demand for greater medical care.

¹If it is assumed that consumers would buy the efficient quantity of health care when they pay market price, then the base of the Harberger triangle measuring the magnitude of the welfare loss due to insurance is exactly the increase in demand caused by insurance (Pauly, 1968).

As a strictly empirical matter, the demand curve for health care certainly slopes down. The Health Insurance Experiment conducted by the Rand Corporation in the 1970s (to draw on the best example of empirical research), examined situations where providers were all paid on a fee-for-service basis, but demanders faced various levels of demand-side cost sharing. Their results clearly demonstrated a large, statistically significant effect of demand-side cost sharing on the level of health services. A plan with a large deductible of up to \$1000 (in the 1970s), for example, reduced total expenditures on health care by 31 percent relative to a plan with full coverage (Manning et al., 1987).

Economists and others concerned with health care policy have debated the meaning of “demand” in health care. After all, as Arrow (1963) emphasized early on, the relationship between patients and their physicians (and other health care providers) involves agency, information, trust, and professionalism. Patients rely on providers to help them articulate their own demand for care. Patients may go beyond this and cede partial or complete authority to providers to make treatment decisions. In health discussions, physicians are sometimes accused of “inducing demand,” which connotes the practice of using influence in the provider’s self-interest. A key unresolved issue in health policy is the magnitude of this provider-induced demand, and how it is determined. In this context, the economists’ notion of a fully-informed and price-taking consumer deciding upon what quantity of health care services to purchase is clearly somewhat off the mark.

However, the optimal insurance literature has been based on the assumption that the demand curve correctly reflects the marginal benefits of services. As a result, it has held that the greater the demand response to what consumer/patients must pay for health care, the higher should be demand-side cost sharing (Zeckhauser, 1970).² As already noted, if demand-side cost sharing is used to reduce consumption, it imposes additional financial risk on consumers.

Supply-Side Incentives

The traditional literature concerning optimal health insurance has been premised on the strong assumption that the quantity of health services is determined purely by the point on the demand curve chosen by the informed, utility-maximizing consumer—after insurance has altered the price the consumer pays. From the individual’s point of view, supply is assumed to accommodate demand fully. This belief reflected the most common method of

²The degree of demand response matters for figuring the optimal demand-side cost sharing even without subscription to the strong interpretation of the demand curve as a marginal benefit schedule. The magnitude of the efficiency loss moving away from any utilization target in response to a reduction in demand-side cost sharing remains directly proportional to the size of the demand response.

payment for health services in the 1960s and '70s, in which hospitals and doctors were paid fees on a per service basis that were set to equal or exceed the costs of each service provided. Under this "fee-for-service" payment system—which continues to be the basis of payment for most physician and hospital care in the United States—health care providers have economic incentives to supply whatever consumption level is desired by their patients.

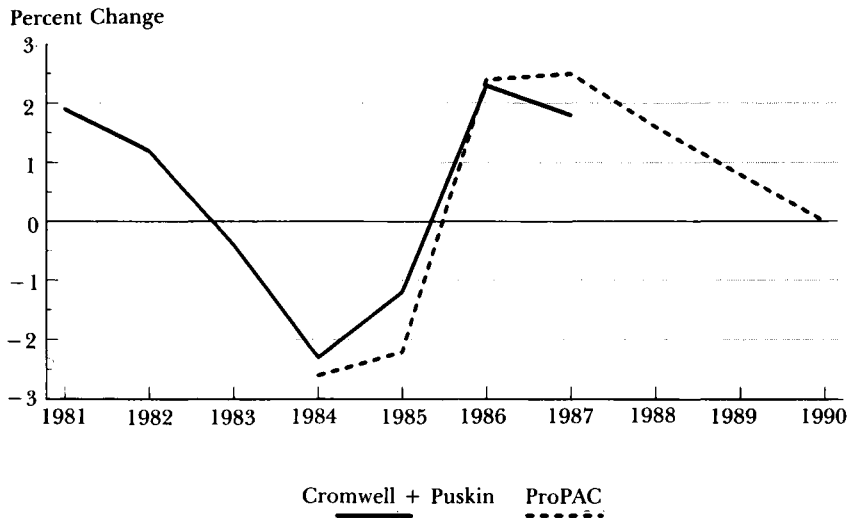
However, as noted earlier, the level of medical care demanded by patients is considerably influenced by health care suppliers. Supply-side cost sharing attempts to alter directly the incentives facing providers. After explaining the idea of supply-side cost sharing and giving some examples of its use in practice, we stress the superiority of supply-side cost sharing in one key respect: changing supply-side cost sharing does not impose financial risk on patients.

Supply-side cost sharing has been a major focus of cost containment over the past decade. Prior to 1982, fee-for-service health care payment systems were intended to cover the cost of care hospitals provided to patients. Payers with substantial market power (the federal Medicare program for the elderly and disabled, state Medicaid programs for the poor and medically needy, and large Blue Cross plans) attempted to pay "only" cost, while commercial insurance companies with small market shares typically paid prices set by hospitals. However, hospital reimbursement changed radically in 1983 with introduction of Medicare's Prospective Payment system (PPS), which fixes payment rates according to the Diagnosis Related Group (DRG) into which the patient is classified at discharge. PPS was based on the idea that hospital "output" can be proxied by the number of discharges in various categories, and that manipulating payment incentives to hospitals could reduce the level and the rate of growth of cost.

Consider, for example, a patient hospitalized for a hip replacement. Under a cost-based payment system, the hospital receives additional funding for each day of care, each diagnostic test, and each procedure conducted. Under the Prospective Payment System, the hospital is given a lump-sum payment equal to the national average of all patients who have a hip replacement (with some adjustments). If the patient is less expensive than the average, the hospital earns a profit on that patient. If the patient remains in the hospital for a long time, or undergoes unusually expensive procedures while in the hospital, then it earns a loss on that patient. Clearly, the move from a cost-based payment system to a prospective payment system dramatically alters the incentives of the hospital to keep the patient longer, to insert a better but more expensive hip, or to perform more expensive procedures. This is because the hospital's revenues per case are now essentially unrelated to the resources expended in treatment of any particular case.³

³The Prospective Payment System does contain provisions which partially compensate hospitals for very high-cost patients. However, since these outlier payments are less than 5 percent of total payments, they perform a fairly minor role in insuring hospitals against losses from expensive patients (Ellis and McGuire, 1988). Hospitals may collect no revenue from patients for Medicare-covered services.

Figure 1

Change in Average Cost of Medicare Patients Discharged from Hospitals

Note: ProPAC data are casemix-adjusted.

The Prospective Payment System was intended to alter incentives to hospitals and improve efficiency. Judith Lave, an economist who helped design the PPS, wrote in 1984 (pp. 260–61): “The per case system should promote efficiency in the production of health care services and in the development and adoption of cost-reducing technologies.”

What has been the impact of the Prospective Payment System on hospital behavior? Consistent with Lave’s predictions, PPS reduced the intensity of resource inputs used (cost) per discharge. Figure 1 draws on two studies spanning two time periods to suggest that PPS had at least a one-time impact on cost growth at short-term general hospitals. The cost per hospital discharge was growing at roughly 2 percent per year prior to PPS, then declined at about 2 percent per year in the periods immediately following implementation. Historical growth resumed three years after the advent of PPS, but has recently declined again.⁴

In economic terms, a prospective payment system can be interpreted as supply-side cost sharing. Once a hospital accepts a patient and qualifies for a DRG-based prospective payment, costs of treatment are borne by the hospital itself; if treatment costs for a patient go up by \$1, the hospital’s net revenue for that patient falls by exactly \$1. A pure prospective payment system can thus be

⁴Other studies support the impact of PPS on hospital costs per discharge. See Hodgkin and McGuire (1993) for a review. Other changes in hospital use during the mid-1980s, apparently in response to PPS, have also been studied, including changes in the number of admissions, transfers between hospitals, the adoption of technology, the actual and reported severity of patients, and the quality of care. See Coulam and Gaumer (1991) for a review.

thought of as complete supply-side cost sharing, and is analogous on the demand side to complete demand-side cost sharing, which is more commonly called “no insurance.” In the no-insurance case on the demand side and the prospective payment case on the supply side, incentives to demand and supply care during the episode, respectively, are minimized by asking the buyer or seller to bear the marginal costs of treatment. Note that because of the intermediary role of the insurer, demand- and supply-side cost sharing can be set independently. It is possible to have any degree of insurance coverage and any degree of supply-side cost sharing. (In particular, there is no sense in which demand- and supply-side cost sharing must “sum to one.”)

Recognizing that prospective payment is a form of supply-side cost sharing leads to other observations and analogies to demand-side cost sharing. The extremes of demand-side cost sharing are complete insurance, and no insurance. The extremes of supply-side cost sharing are pure prospective payment, and pure cost-based reimbursement. Of course, intermediate choices are available; there is a family of insurance plans characterized by the coinsurance, and a family of supply-side payment systems characterized by the share of costs at the margin borne by the provider.⁵

Once prospective payment is accepted as a member of a family of payment systems (and an extreme member at that) it is possible to ask questions about the “optimal degree of supply-side cost sharing,” just as this question has been asked about demand-side cost sharing. It is by no means self-evident that the best pick is a purely prospective system.⁶ After all, the supply of care will respond to the degree of supply-side cost sharing in two potentially negative ways. First, providers in such a system will tend to seek out patients with a high expected profit, while deterring patients expected to generate losses (Dranove, 1987). In addition, given the ability of health care suppliers to shape the amount of care provided, complete supply-side cost sharing may lead to some underprovision of care.

Why is it likely that the quantity of services supplied to an individual seeking treatment will tend to rise as the degree of supply-side cost sharing falls, and as the system moves away from prospective payment and toward cost-based reimbursement? This relationship can be interpreted as a sort of supply curve, and one rationale for its upward-slope is based on the tradeoff a

⁵Formally, this family of supply-side payment systems can be characterized by a prospective payment amount “ R ” independent of costs, and a share “ r ” of costs paid. The share of provider costs not directly reimbursed, what we call supply-side cost sharing, is $s = 1 - r$. If costs are x , the payment to a hospital per discharge would be $R + (1 - s)x$. When $R > 0$ and $r = 0$, then $s = 1$ and the system is completely prospective. When $R = 0$ and $r = 1$, then $s = 0$ and the system is cost-based reimbursement. “Mixed systems,” partly prospective and partly cost-based are when $R > 0$, and $0 < s < 1$. Demand-side cost sharing can be modeled in an analogous fashion, with a portion of costs paid for by premiums, π , and a portion c paid for when services are consumed.

⁶We have examined this question within an explicit normative framework close to the conventional theory of optimal insurance in Ellis and McGuire (1986, 1990). We make the case for supply-side cost sharing more generally here.

payment system imposes on the clinical decision-maker.⁷ Ellis and McGuire (1986, 1990) suppose that each health care provider is an agent acting on behalf of the imperfectly informed patient. As long as the benefit to the patient has some weight in the objective function of the health care provider, then increasing the marginal reimbursement to providers will increase the desired level of services supplied, since additional benefits to the patient (which also provides utility to the provider) can be purchased at lower cost.⁸

Several real-world reimbursement systems employ intermediate levels of supply-side cost sharing. The most important U.S. example is the reimbursement formula used by Medicare to reimburse specialty hospitals (for example, most psychiatric, rehabilitation, and children's hospitals) for their 300,000 discharges each year, which are paid under the TEFRA system. TEFRA, an acronym for the Tax Equity and Fiscal Responsibility Act of 1982, set up a payment system based around a "target cost" at each hospital. If actual cost per discharge exceeds or falls below the target by specified amounts, Medicare pays part of the difference, a form of supply-side cost sharing (Cromwell et al., 1992). Health maintenance organizations (HMOs) also provide examples of supply-side cost sharing. Most HMOs accept a fixed annual capitation payment for each enrollee, and are then responsible for all costs that enrollee may incur during the year. Many HMOs then compensate their physicians using a financial incentive system that include both a salary payment and a payment based on the volume of services provided. Several European countries are experimenting with mixed reimbursement systems. For example, Danish physicians are paid using a formula under which 75 percent of their earnings are from fee-for-service and 25 percent from lump-sum payments based on the total number of enrollees under their supervision ("capitation" payments). This new payment formula, together with their budgeted payments to hospitals, appears to have helped Denmark control its health costs (Abel-Smith, 1992).

There is considerable evidence that payers can affect utilization through their choice of the degree of supply-side cost sharing. If it is presumed that too much health care results when a hospital is paid on a fee-for-service basis to treat fully insured patients, an increase in the degree of supply-side cost sharing can be imposed to reduce utilization. Moreover, supply-side cost sharing can reduce utilization without shifting costs to patients and discarding the risk-spreading objectives addressed by health insurance.

⁷Decisions about resources use in hospitals are made jointly by many agents, including hospital staff and doctors. Hospital supply-side payments do not directly impact doctors financially, but the literature documenting the fall in patient's length of stay following introduction of prospective payment suggests that doctors do respond to the interests of the hospital staff.

⁸Pope (1989) also offers an argument for why less supply-side cost sharing leads to a greater quantity of care provided, using a model of hospital nonprice competition. In his model, hospitals compete to attract patients through offering quality, which is modeled as a public good available to all patients at a hospital. In markets where nonprice competition is weak, a fully prospective payment system will tend to be too little quality. Supply-side cost sharing is good in his model, because by subsidizing quality, it encourages more to be provided.

A Framework for Combining Demand-Side and Supply-Side Cost Sharing

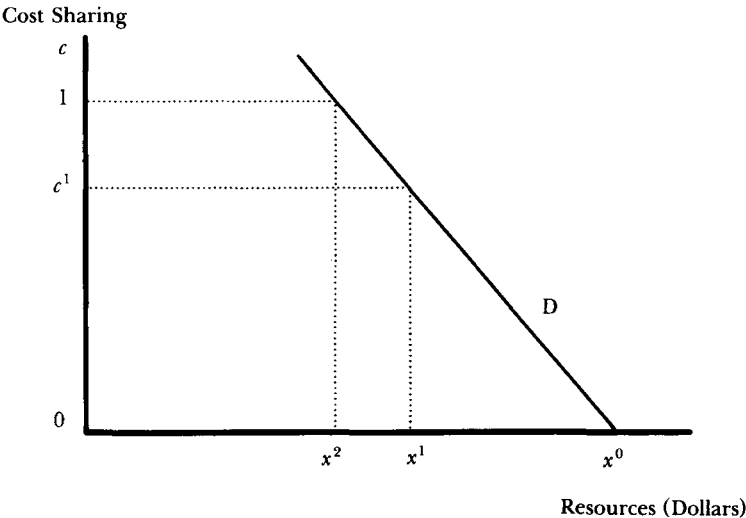
As an illustration of how demand-side and supply-side cost sharing would function in the market, consider the hip replacement episode mentioned already. Most employer-based health insurance plans provide full or nearly full coverage for hospital expenses, so patients who are covered by such plans would pay nothing or a relatively small fraction of costs of treatment in the hospital. Figure 2 depicts a demand curve of one patient for care during the hospital stay for hip replacement. On the vertical axis is shown the price, expressed in terms of cost sharing, the percentage of costs paid for by the patient (c). The horizontal axis measures the resources devoted to the patient's care during the stay, reflecting the reality that hip replacements can be performed in a number of different ways. These resources can be measured in dollar terms, by using some fixed set of prices to aggregate the different possible services.⁹ When the patient is fully insured, facing no price for care at the time service is provided, the patient would demand quantity x^0 .

If the consumer were fully informed and could choose the level of care to maximize utility, standard consumer demand theory dictates that quantity x^2 be labelled the socially efficient level of care. This is the quantity the fully informed consumer would buy facing the social cost of consumption. However, since we are skeptical that the observed demand can be interpreted as reflecting "socially efficient" consumption, we interpret the demand curve in a more limited way, as an empirical relationship between degree of cost sharing and quantity of use demanded by the patient. To be explicit, this demand curve is drawn while holding constant all other prices, including prices paid to the hospital for care. For concreteness, assume that the demand curve describes the resources used for the hip replacement episode as demand-side cost sharing changes, given that the hospital has no supply-side cost sharing. We arbitrarily designate the socially desired quantity of treatment to be x^1 . (This x^1 need not be more than x^2 for our argument to hold.) As the demand curve describes, this quantity could be achieved by a coinsurance of c^1 , but this would entail substantial financial risk to consumers. If a hip replacement were needed, the out-of-pocket costs to the patient would be c^1x^1 .

Now turn to Figure 3, where the supply curve of care to this patient is drawn. This supply curve indicates how the quantity of care supplied during the episode responds to the degree of supply-side cost sharing. Here, we measure on the vertical axis the reimbursement price r , the share of costs reimbursed by the insurer. (Note that supply-side cost sharing, $s = 1 - r$, increases as we move down the vertical axis from one toward zero.) Again, we interpret this supply curve only as an empirical relationship, and do not

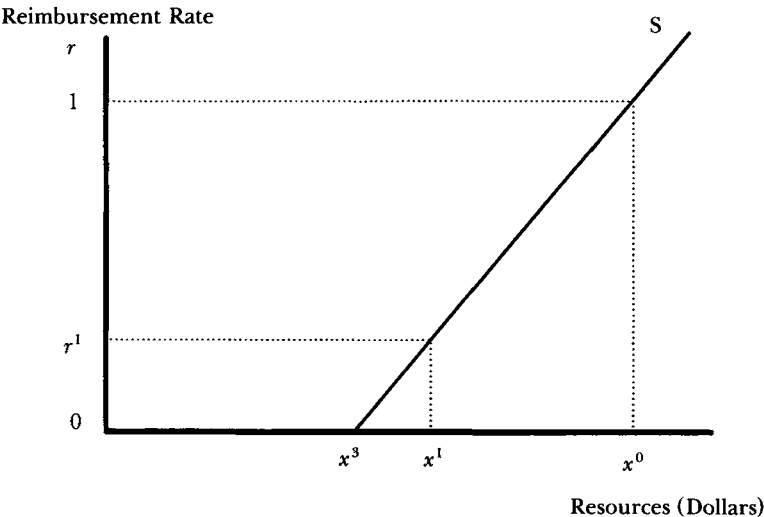
⁹One rationale for this approach is that price (the cost-share) multiplied by the quantity (expenditures) yields total out-of-pocket costs, as usual.

Figure 2
Demand Curve for a Hip Replacement



attribute special normative meaning to levels of supply corresponding to cost-based reimbursement or prospective payment. Similarly to the demand curve, this supply is drawn assuming no demand-side cost sharing; or to put it another way, this supply curve describes utilization at the hospital contingent on serving fully-insured patients, corresponding closely to the form of most empirical research on hospital supply.

Figure 3
Supply Curve for a Hip Replacement



The levels of resources labeled x^0 in both Figures 2 and 3 refer to the amount of utilization with no demand- or supply-side cost sharing—and thus x^0 must be the same amount in both diagrams. A fully prospective payment system with $r = 0$ and hence supply-side cost sharing equal to one (and demand-side cost sharing still at zero) would lead to x^3 . The desired quantity x^1 , chosen to be the same as in the earlier demand curve, could be implemented by a degree of supply-side cost sharing equal to r^1 , with no demand-side cost sharing.¹⁰

With both demand- and supply-side cost sharing available as payment parameters, the health payment system can do a better job in promoting both social goals of risk protection and efficient health care provision. In fact, a natural division of labor emerges between the two sets of instruments. Where possible, use supply-side cost sharing to achieve health care utilization targets, such as a desired cost per episode of inpatient care. Use demand-side cost sharing to protect patients against financial risk. The use of both tools is likely to come closer to the first-best outcome of minimizing consumer's financial risk while offsetting the moral hazard problem. Depending upon the provider's objective function, and the nature of costs, technology, demand, and competition, in some cases a "first best" can be achieved using full insurance to protect the consumer from financial risk, and partial supply-side cost sharing—something we have labeled a mixed system—to constrain demand.

However, it may be that even pure prospective payments will lead to the provision of a level of services that exceeds the socially optimal level. In terms of Figures 2 and 3, this is the same as noting that without additional assumptions, we cannot say whether X^1 is larger or smaller than X^3 . An example of this situation would be in a health maintenance organization where patients may tend to make too many outpatient visits even when the provider has strong incentives to discourage such excess use. In such circumstances, demand-side cost sharing incentives may still be desirable.

An important insight from synthesizing supply- and demand-side payment incentives is the recognition that patients and providers need not agree on the level of services to be provided. It may well be, for example, that fully insured patients demand more services than providers paid under some system with supply-side cost sharing will agree to give them. (Readers enrolled in health

¹⁰More formally, we could characterize the quantity of services provided to this patient as $X = X^*(c, r, \pi, R, Z)$ where c and r are the demand-side cost share and supply-side reimbursement rate, respectively, π is the consumer's health premium, R is the lump sum payment paid to the provider, and Z is the complete set of all other demand- and supply-side variables that influence the level of services provided. If we restrict ourselves to actuarially fair premiums and lump sum payments that just cover costs (meaning that π and R are each chosen so that payments equal expenditures), then this imposes constraints on π and R , which could be used to rewrite X as $X = X(c, r)$. Figure 2 plots X as c is changed from 0 to 1, holding $r = 1$. Figure 3 plots X as r changes from 0 to 1, holding $c = 0$. Note that $X^0 = X(0, 1)$, $X^1 = X(c^1, 1) = X(0, r^1)$, $X^2 = X(1, 1)$, and $X^3 = X(0, 0)$. Our conclusions do not depend on the underlying model generating the $X = X(\cdot)$ function, only the fact that a target can be hit by either changing c or changing r .

care systems that make active use of supply-side cost sharing, such as prepaid group practices, will be familiar with this idea.) Furthermore, although it may appear contradictory at first glance, consumers may like it this way in the sense that a rational consumer, if given a choice, may select into a payment system where patients can be expected to be restrained in health care use by providers.

Other Goals for Supply-Side Cost Sharing

The preceding discussion has emphasized that supply-side cost sharing can be used to hit utilization targets without imposing risk on patients. In this section we briefly discuss other arguments in support of supply-side cost sharing: appropriate incentives towards new technologies, fairness to health care providers, and fairness to consumers.

New Technologies

Thus far, the discussion has focused on the supply and demand determinants of care to a single patient, in a static context. In the medium to long term, the development and adoption of new technologies is an extremely important determinant of costs and service patterns (Newhouse, 1992). A growing number of writers have highlighted the association of insurance- and cost-based reimbursement with new technologies, and thus with the rapid rate of increase of health care costs (Weisbrod, 1991; Baumgardner, 1991).

As one example, U.S. hospital care is more intense than in the past, and more intense than other countries. In 1993, the average cost in a general hospital will exceed \$1,000 per day.¹¹ The real increase in cost per day has gone up by a factor of five since 1960. Application of more sophisticated medicine, with implications for both capital and labor costs, is usually cited as the proximate cause.¹² As Arnold Relman, long-time editor (now retired) of the *New England Journal of Medicine* bluntly put it, technology is the “engine behind the rise in medical costs,” fueled by the growing number of doctors “trained to provide high-tech, expensive services” (Relman, 1989).

¹¹The data in Fisher (1992, pp. 4–5) for 1989 show that the average cost per day in 1989 was already about \$900.

¹²In a recent article in this journal, Joseph Newhouse (1992) examined health care cost containment in a dynamic setting, trying to explain the rapid growth in health costs over the past 50 years. Although aging, increasing insurance coverage, increased income, supply-induced demand, malpractice, slow factor productivity growth, tax subsidies to health care, and the high costs of treating the terminally ill were each identified as partial explanations for the increases, his bottom line conclusion is that these factors together “account for well under half—perhaps under a quarter—of the 50-year increase in medical care expenditures” (p. 11). He attributes the bulk of the increase to cost-increasing technological change. Most of the examples given by Newhouse involve hospital treatment, thus explaining why costs of the hospital sector have grown faster than most other sectors over the past five decades.

While growth in costs and the technological sophistication is not obviously bad (indeed, such growth can be cited as a major strength of the U.S. health care delivery system), many observers pronounce unfavorably on the course taken by the U.S. health care sector. Comparing the American and Canadian experience, Robert Evans (1986, p. 602) concludes, "It is notorious that new interventions in the U.S. have tended to proliferate too rapidly, far in advance of their evaluation, so that patients are exposed to unproven technologies which may turn out to be of little value or even harmful in some or all of their applications." In his review of U.S. health policy and diffusion of innovation, Bruce Hillman (1986, p. 683) argues that the American-specific "elements of medical regulation, reimbursement and competition spur unwarranted diffusion of medical innovations before their utility is understood..." Generous insurance coverage increases patients' demand for quality-enhancing treatment innovation, while making patients indifferent to the cost.

We concur with this assessment of the key role of technological change in explaining cost increases in the United States. The combination of imperfect information, generous insurance coverage, and cost-based reimbursement of providers creates strong incentives, probably too strong, for health care providers to compete by purchasing and using new, expensive technologies. However, demand- or supply-side cost sharing can affect the rate of growth of costs, as well as the level of costs per person or per episode of care. If society's goal is to hit a particular cost/technology-growth target, supply-side cost sharing continues to have the advantage that it does not force patients to bear financial risk. Furthermore, given the informational advantage that providers have over consumers in evaluating the effectiveness of new medical technologies supply-side incentives would appear more desirable than demand-side methods as a means of affecting the pace of new technology adoption.

Reducing Risk and Increasing Fairness in Payment to Providers

A system that is fair to providers of health care should presumably have a minimal difference between expected payments and costs. However, neither a purely cost-based nor a prospective payment system is likely to be the most fair, in the sense of paying for true costs of each patient, as Pope (1990) demonstrates. The fundamental problem here is that payers for health services can only imperfectly observe the true costs borne by health care providers.¹³

Use of estimated costs in a payment system can improve fairness in relation to complete prospective payment. In Ellis and McGuire (1988), we explore the fairness of the payment system to providers with a focus on the biased selection

¹³This argument is closely related to analyses in the broader principal-agent literature on contracting (for a starting point in this journal, see Sappington, 1991). In that literature, the principal (corresponding to our insurer) writes a contract with an agent (the hospital) to produce output (care for a patient). The principal cannot observe patient type (severity) or directly specify the resources the hospital uses in treatment. The second-best contract generally features cost-sharing between the insurer and the hospital, which corresponds to supply-side cost sharing.

which results when prospectively paid providers seek to attract a low-cost group of patients. We demonstrate that an intermediate level of supply-side cost sharing can substantially reduce the financial gain to insurers from attracting favorable selections of patients, or the loss from being stuck with an unfavorable group. Such a system reduces the risk a provider bears because payments are partly related to costs, and the expected deviation of payments and costs is reduced in comparison to a pure prospective payment system. Siegel et al. (1991) use an empirical Bayesian approach to demonstrate that given that hospital costs attributable to a given patient are only imperfectly observed, the fairest reimbursement system is one that uses a weighted average of the patient's own cost, the average cost of other patients at the hospital, and some average using other hospitals (such as a regional or national average).¹⁴

Each of these models can be interpreted as supporting intermediate levels of supply-side cost sharing, or what we have labeled a mixed system. In one form or another, a mixed system can be more fair to providers than either extreme.

Fairness to Consumers and Access to Health Care

Access to health care for all is a fairness goal that can be in conflict with provision of efficient health care services. In practical terms, demand-side cost sharing may do most to discourage care among the lower income groups in society. If equity of access is accorded high weight, then supply-side cost sharing will be superior to demand-side cost sharing as a tool for cost containment.

Many other countries have already made this decision. Very little reliance upon demand-side cost sharing is made in most European countries (Schieber et al., 1992). Instead, European countries have relied upon a wide range of supply-side incentives over the past decade: capitation, salaried employment, expenditure caps, global budgeting, and health planning. All of these countries have had less of a cost escalation problem than the United States in recent years. Schieber et al. make the important observation that the United States and Canada, the two countries that rely most extensively upon fee-for-service reimbursement for physician services, also had the most rapid expenditure growth of health care as a percentage of national income.

A Policy of Supply-Side Cost Control in Health Care

The federal government appears ready to regulate the form of demand-side cost sharing in the United States by guaranteeing universal insurance and

¹⁴In the terminology of hospital payment, Siegel et al. (1991) advocate a system with a "blend" and a "mix." The blend refers to calculating the fixed component of payment as a combination of the facility's own historical average cost and a national average. The mix element refers to payment based partly on the cost of the actual case.

specifying a minimum amount of coverage everyone must have, thus limiting the amount of demand-side cost sharing that individuals will be allowed to accept. Also likely to be proposed is the idea of creating local insuring organizations, sometimes referred to as Accountable Health Plans (AHPs), with both the incentive to bargain hard with hospitals and other providers (because their revenue would be predetermined premiums) and the power to extract good terms (because only a few AHPs would be operating in each region). Implicit in the discussion of AHPs is the assumption that many of these AHPs will use supply-side incentives to change provider behavior and contain costs—but the actual levels of supply-side cost sharing will be left up to the AHPs. Should the federal government be considering more dramatic supply-side policies?

One extreme supply-side policy would be for the federal government to nationalize health insurance and pay providers prospectively or use some intermediate form of supply-side cost sharing. Legislation that would implement a Canadian style national health insurance system, and also implement supply-side cost containment incentives appears to have support from a significant minority of congressional delegates (with over 80 sponsors as of July 1993), but strong opposition from insurers and many provider and consumer groups make this extreme form unlikely to be implemented in the near future.

A somewhat less extreme approach for implementing supply-side cost sharing is for the federal government to coordinate payers' policy, requiring all insurers to pay providers using the same formulas. Such "all-payer" systems, which have been tried in several states and are popular in many European countries, could be used to change dramatically incentives and the nature of competition between providers, and likely result in major reorganizations of health care delivery. If an all-payer system is deemed desirable, it seems attractive to let individual states (or smaller areas) innovate and try out different payment systems in the post-health reform environment before imposing a uniform system at the national level.

A more moderate supply-side policy is for the federal government to encourage the private sector to adopt sounder supply-side payment policies. Two specific approaches for the federal government to do this are to change the incentives for private payers to use supply-side cost sharing, and to set a good example in its own payment formulas.

One important way that the government can change incentives to adopt supply-side cost sharing is through the tax system. Much has been written by economists about the desirability of limiting the tax deductibility of both employer and employee contributions to health insurance premiums (Pauly, 1986; Enthoven and Kronick, 1991). The usual argument has focused on the subsidy leading to "overinsurance" on the demand side. By a similar argument, the tax subsidy to health care premiums encourages overinsurance on the supply side—encouraging consumers to select too little supply-side cost sharing. Reducing the tax subsidy is likely to encourage more people to choose health maintenance organizations, preferred-provider organizations, and other networks that use supply-side incentives.

A second way that the government can encourage supply-side cost sharing is by giving the private sector better tools for doing so. For example, the government plays an important role in facilitating private choices of supply-side cost sharing by developing ways of defining and classifying health care "products." These might be related to the DRG classification system but with more flexible terms. Each type of product (whether a discharge, an outpatient episode, or other group of services) may need a relative value, with appropriate case-mix or geographic adjusters. Peer groupings of hospitals or other providers serving distinct types of patients may be necessary to reflect desirable cost differences. Private payers can then be encouraged to use this product-definition system to negotiate forms of payment with providers. Of course, they might use the definitions in different ways: private payer 1 might use a completely prospective system; private payer 2 might pay using a mixed system and contract only with a subset of hospitals; and so on. If private payers are going to be relied upon to control costs, then it is the responsibility of the public sector to give them better tools.

Along with providing better tools, the government should set a good example of using those tools with its own methods of supply-side cost sharing. In 1990, the federal Medicare program accounted for 33 percent of hospital revenues. Adding state-operated Medicaid programs and other government programs brings the public share of hospital revenue to nearly 50 percent (ProPAC, 1992, p. 27). These public programs have made heavy use of supply-side cost sharing, particularly for hospital payments. Medicare pays hospitals by DRGs, many Medicaid programs use DRGs, others have generally low payment levels (ProPAC, 1992, pp. 27-28). In contrast, only about one-third of private payers use some form of supply-side cost sharing for hospital payment. More common are payments on the basis of charges, discounted charges, or negotiated prices (ProPAC, 1992).

Why have private payers been slow to follow Medicare and some Medicaid programs and adopt a prospective payment system? This question is crucial for considering what public policy should be towards payment systems elected by private payers. One reason may be that supply-side cost sharing in the public sectors has gone too far. Medicare and Medicaid, the public payers making the most aggressive use of supply-side cost sharing, serve captive populations. If you are old (and in Medicare) or poor (and in Medicaid), you have no choice about the hospital payment system in which you participate. Private plans, where employees and their families do have choices, have not embraced the federal DRG-based Prospective Payment System. Incentives to restrain use in a fully prospective system may be too great, and a mixed system should perhaps be tried.

The self-interest of the public sector may be a real impediment to development of a fair and efficient supply-side cost sharing policy. Government payers can be aggressive in setting payment terms because they can credibly commit to tough payment rules and force private payers to pick up more than their share of hospital joint costs (Ma and McGuire, forthcoming 1993). Data from ProPAC

(1992) shows that the state Medicaid programs pay on average only about 80 percent of their allocated costs, and the federal Medicare program pays about 90 percent. This finding reflects a practice known as "cost shifting." Furthermore, hospitals are able to provide care to the uninsured in large part because of the high price/cost margins they charge private payers. Governments do not simply set the rules of the game in health care, they are big players themselves. In the context of health care reform, it would be unfortunate if the federal government concentrates on keeping health costs down for Medicare only, rather than for the country as a whole, and is tempted to let the private sector continue to fend for itself in its economic relationship with health-care providers.

■ *Research for this paper was partially supported by grant K05-MH00832 from the National Institute of Mental Health. The authors are grateful to Alan Krueger, Joseph Newhouse, Carl Shapiro, and Timothy Taylor for their detailed and useful comments.*

References

- Abel-Smith, Brian, "Cost Containment and New Priorities in the European Community," *Milbank Quarterly*, 1992, 70:3, 393-416.
- Arrow, Kenneth, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, December 1963, 53, 941-73.
- Baumgardner, James R., "The Interaction Between Forms of Insurance Contract and Types of Technological Change in Medical Care," *Rand Journal of Economics*, Spring 1991, 22:1, 36-53.
- Burner, Sally T., Daniel R. Waldo, and David R. McKusick, "National Health Expenditures Projections Through 2030," *Health Care Financing Review*, Fall 1992, 14:1, 1-29.
- Coulam, Robert F., and Gary L. Gaumer, "Medicare's Prospective Payment System: A Critical Appraisal," *Health Care Financing Review*, Annual Supplement, 1991, 45-77.
- Cromwell, Jerry, and Dena Puskin, "Hospital Productivity and Intensity Trends," *Inquiry*, Fall 1989, 26:3, 366-80.
- Cromwell, Jerry, Randall P. Ellis, Brooke Harrow, and Thomas G. McGuire, "A Modified TEFRA System for Medicare Discharges from Psychiatric Facilities." In Frank, R., and W. Manning, eds., *Economics of Mental Health: Essays in Honor of Carl Taube*. Baltimore: Johns Hopkins University Press, 1992, 41-67.
- Dranove, David, "Rate-Setting by Diagnosis Related Groups and Hospital Specialization," *Rand Journal of Economics*, Autumn 1987, 18:3, 417-27.
- Ellis, Randall P., and Thomas G. McGuire, "Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, Summer 1986, 5:2, 129-52.
- Ellis, Randall P., and Thomas G. McGuire, "Insurance Principles and the Design of Prospective Payment Systems," *Journal of Health Economics*, September 1988, 7:3, 215-37.
- Ellis, Randall P., and Thomas G. McGuire, "Optimal Payment Systems for Health Services," *Journal of Health Economics*, December 1990, 9:4, 375-96.
- Enthoven, Alain, and R. Kronick, "Universal Health Insurance Through Incentives Reform," *Journal of the American Medical Association*, May 15, 1991, 265:19, 1532-37.
- Evans, Robert G., "Finding the Levers, Finding the Courage: Lessons from Cost

Containment in North America," *Journal of Health Politics, Policy and Law*, 1986, 11:4, 585-616.

Fisher, Charles R., "Trends in Total Hospital Financial Performance Under the Prospective Payment System," *Health Care Financing Review*, Spring 1992, 13:3, 1-16.

Hillman, Bruce, "Government Health Policy and the Diffusion of New Medical Devices," *Health Services Research*, December 1986, 21:5, 681-711.

Hodgkin, Dominic, and Thomas G. McGuire, "Payment Levels and Hospital Response to Prospective Reimbursement," unpublished manuscript, Boston University, 1993.

Lave, Judith R., "Hospital Reimbursement Under Medicare," *Milbank Memorial Fund Quarterly*, 1984, 62:2, 251-67.

Ma, Albert, and Thomas G. McGuire, "Paying for Capital Costs in Health Care," *Journal of Economics and Management Strategy*, forthcoming 1993.

Manning, Willard G., et al., "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review*, June 1987, 77:3, 251-77.

Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives*, Summer 1992, 6:3, 3-21.

Pauly, Mark, "The Economics of Moral Hazard: Comment," *American Economic Review*, June 1968, 58:3, 531-37.

Pauly, Mark, "Taxation, Health Insurance and Market Failure in the Medical Economy," *Journal of Economic Literature*, June 1986, 24:2, 629-75.

Pope, Gregory C., "Hospital Nonprice Competition and Medicare Reimbursement Policy," *Journal of Health Economics*, June 1989, 8:2, 147-72.

Pope, Gregory C., "Using Hospital-Specific Costs to Improve the Fairness of Prospective Reimbursement," *Journal of Health Economics*, November 1990, 9:3, 237-51.

Prospective Payment Assessment Commission (ProPAC), *Medicare and the American Health Care System: Report to Congress*, Washington, D.C., June 1992.

Reiman, Arnold, "Confronting the Crisis in Health Care," *Technology Review*, July 1989, 92:5, 30-40.

Sappington, David E. M., "Incentives in Principal-Agent Relationships," *The Journal of Economic Perspectives*, Spring 1991, 5:2, 45-66.

Schieber, George, Jean-Pierre Poullier, and Leslie M. Greenwald, "U.S. Health Expenditure Performance: An International Comparison and Data Update," *Health Care Financing Review*, Summer 1992, 13:4, 1-88.

Siegel, Carol, et al., "A Risk Based Prospective Payment System that Integrates Patient, Hospital, and National Costs," *Journal of Health Economics*, 1991, 11:1, 1-42.

Weisbrod, Burton, "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment," *Journal of Economic Literature*, June 1991, 29:2, 523-52.

Zeckhauser, Richard, "Medical Insurance: A Case Study of the Tradeoff Between Risk Spreading and Appropriate Incentives," *Journal of Economic Theory*, March 1970, 2:1, 10-26.

