Health Care Cost Containment: Some Implications of Global Budgets

Carl M. Stevens

There have been in recent years calls from many quarters for reform of the United States health care system. Indeed, what to do about the nation's health care system became a major issue in 1992 political campaigns at both state and federal levels. The list of alleged malfunctions is long. For example, it is contended that health care claims too large a share of the nation's total output of goods and services, that access to services is unfairly limited, including the problem of the uninsured, that there is too much unnecessary medicine owing in part to fear of malpractice suits, and that there is too little knowledge about the health impact of many common medical procedures, among other problems.

All of these problems warrant and are getting substantial attention. However, most widespread and urgent is the call for cost containment and other reforms of the health care financing system, including those that would address equity of access. As U. Reinhardt has remarked, "People do not wish to reform their doctor's office or hospital, they want the financial arrangement restructured" (*1*).

Given such priorities at the household level, it is important to understand that it may not be possible to restructure the financial arrangements in ways that would satisfy cost containment without at the same time having a substantial impact on the health services delivery system—that is, without, as a by-product, reforming the doctor's office and hospital. There are no nontrivial changes in health financing systems that do not affect the performance of the health care delivery system. Successful reform of the health financing system may be impeded by various factors—for example, vested interests in existing arrangements and others. Important among these others is widespread public misunderstanding of the terms frequently used by parties in the health reform debate and, more generally, failure to appreciate the implications of commonly suggested financing reforms for the performance of the delivery system.

There is thus a real need to promote public understanding in this domain. Here, in drawing attention to some of the implications of changes in health financing arrangements for the delivery system, I focus on global budgets, which are frequently called "expenditure caps" or just "caps"—that is, administratively set limits on overall health care spending. They are the most potent weapon in the arsenal of cost containment strategies and they are increasingly being urged as the strategy of choice. Advocacy of the Canadian health care financing system implicitly represents an endorsement of such caps. As time passes and if (as is likely) alternative, less Draconian cost containment strategies fail to achieve substantial reductions in the rate of increase in health care spending, the call for expenditure caps will grow in intensity. Vermont has enacted legislation that, if implemented, will make it the only state so far to implement a global health budget, capping both public and private medical spending (*2*). Also, a bill providing a Canadian-type health care financing system was recently sponsored by an Ohio state legislator. Under this plan, all private health insurance would be phased out in 3 years. The Ohio Health Care Trust Fund financed by various taxes (and incorporating Medicare and Medicaid funding) would be created as the single payer, administered by the Ohio Health Plan Board. The board would negotiate global budgets with each hospital and fees with representatives of physicians (*3*).

Global budgets (expenditure caps) are among the least generally understood of cost containment strategies. This is so because some of the most important implications of this strategy for the performance of the delivery system are less apt to be self-evident than those of other commonly suggested cost containment strategies. The major implications I discuss are: (i) problems in reflecting or representing consumer preferences in the decisions that allocate resources for health care and (ii) implications for providers and consumers (patients) of the non–price rationing necessarily accompanying the imposition of effective expenditure caps. These implications in particular tend to be overlooked or underemphasized in discussions of the power of single-source financing systems to impose global budgets (*4*).

Global Budgets

Global budgeting approaches for physician and hospital services are common in other Western, industrialized countries such as Canada and the United Kingdom (*5*). Spending for these categories of services in 1990 comprised about two-thirds of personal health care expenditures in the United States (*6*). Global budgets entail administrative and management controls to ensure staying within the budget. Expenditure targets, on the other hand, are more in the nature of policy goals rather than an absolute limit on spending.

The "global" in global budgets distinguishes them from special purpose or partial expenditure caps. A global budget is supposed to limit total spending for health services or at least come close to this coverage. For example, in the United Kingdom, almost all consumers use National Health Service (NHS)–budgeted physician and hospital services. The number of patients who go to private physicians and hospitals (including the private wings of NHS hospitals) and who buy private medical insurance is currently estimated at about 5.5 million or about only 8.0% of the population (*7*).

Although a global budgeting approach to cost containment in the U.S. health care system has not been generally regarded as appropriate, there has been serious discussion, and some implementation, of partial expenditure caps, notably for the big federal government health insurance programs Medicare and Medicaid. Historically, these programs have been regarded as entitlement programs under which defined categories of beneficiaries are entitled to a broad range of medical services. Over the years, the budgets for these programs have grown rapidly in response to the increasing use by beneficiaries of their services. This increasing claim on fiscal capacity made by these programs, at both state and federal levels, has resulted in a vigorous search for cost containment strategies. As matters stand, under the diagnostic related groups (DRG) scheme for hospital reimbursement and the resource-based relative value scale (RBRVS) scheme for physician reimbursement, Medicare is attempting to contain costs by implementing price controls across the board. Recently, there has been serious discussion of adopting binding expenditure caps for Medicare.

In the case of Medicaid, a program financed by both federal and state funds for defined categories of the medically indigent, at least one state, Oregon, has proposed an expenditure cap. This has been accompanied by a formal rationing scheme under which the services to be provided to Medicaid beneficiaries have been listed in order of priority from 1 to 703. The services actually to be provided to Medicaid beneficiaries will depend on the budget for this program provided by the state legislature (and the federal matching funds this will

The author is an emeritus professor of economics at Reed College, Portland, OR 97202.

entitle the state to receive). Currently, the program budget will not permit payment for services below 587 on the list (8).

Expenditure caps just for the beneficiaries of public health insurance programs raise rather different public policy issues than global budgeting does. For the former, a central policy question is what we, in our role as relatively affluent, tax-paying citizens, owe our unfortunate fellow citizens who find themselves medically indigent. Such distributional questions are not central to the evaluation of global budgets, which are, after all, supposed to fund the services available to virtually all citizens, not just to those who are welfare beneficiaries of special programs. It is generally accepted that a few citizens at the upper end of the income distribution will be able to pay for and use more services than those provided to the population as a whole under the global budget.

If effective, binding expenditure caps could be implemented for Medicare and Medicaid, this might relieve much of the pressure these programs put on government budgets at the state and federal levels. But such constraints would not achieve the objectives of many proponents of cost containment. In 1991, Medicare and Medicaid together accounted for 29% of total U.S. expenditures for health. Other government programs accounted for another 14%. This leaves the privately funded U.S. health economy accounting for some 59% of total health spending, a sector for which there are also urgent calls for cost containment. Also, under caps for government programs only, there would be large-scale shifting of costs to the privately funded health economy as beneficiaries of public health programs would seek to obtain from the privately funded sector services not available from public programs and as the privately funded sector would find it difficult to rebuff such requests for services.

Global Budgets: How to Implement Them and at What Level?

In much of Europe and Canada where global budgets are implemented, the flow of money from third-party payers to providers comes from what is effectively one large funding source. These single-source health financing systems are central to containing aggregate expenditures for health—that is, to putting limits on the global budgets (4). To the extent that the single source is general tax revenue, direct control over health care expenditures can be imposed by the public budget-making process. There may in addition be negotiations between the funding source and the providers (hospital and physicians) to determine the size of the global budget and other matters. In these negotiations, the single-source payer will wield a large amount of bargaining power.

In the United States, on the other hand, the flow of money to providers comes not from one or a few large sources, but from literally thousands of uncoordinated sources, including patients themselves, business firms, insurance carriers of many different kinds and sizes, all levels of government, and others (9). Presumably, the implementation of global budgets in the United States would require a change in these institutional arrangements to something of the kind that operates in Europe (for example, perhaps one National Health Insurance scheme that takes the place of the existing array of funding sources).

In terms of the implications for the performance of the delivery system, the level at which a global budget might be set is obviously a critical matter. To be effective, a global budget must restrict the flow of resources for health services to less than the resources that would otherwise have been available because otherwise it would not be containing costs. But how much less? Recently, health expenditures in the United States have been at about 12.0% of the gross national product (GNP), the nation's total output of goods and services. By the year 2000, health care expenditures are predicted to be about 16% of the GNP (7). For Canada, Sweden, and Germany, health expenditures have been running around 9.0% of their GNPs. A global budget that would attempt to force health spending in the United States down to something like 9.0% of the U.S. GNP would entail a no doubt unacceptably Draconian constraint. Proponents of global budgets for the United States have been vague about the limits on total health spending they propose. I assume a global budget that would impose a significant constraint (that is, would provide resources for health significantly less than would otherwise have been provided).

Implications for the Delivery System

Whatever the nature of the health financing system, all of us in the aggregate must necessarily pay the economic cost of the nation's health care system and also must, in one way or another, pick up the tab for financing that system. The economic cost is represented by all of the other nonhealth goods and services that could have been produced had the resources committed to the production of health services been used instead to produce other goods and services. In our market-type economic system, consumer preferences are supposed to decide such resource allocation questions. By making choices of what to buy and what not to buy, individuals in their roles as consumers make their preferences known to producers, who then respond to them.

Consequently, to set the appropriate limit on total health spending (that is, on the quantity of health services), the budget authorities (or otherwise constituted planners) must weigh the relative importance to consumers of health services and the other goods and services these same resources might alternatively produce. But where is the required information about consumer preferences to come from? The administrative planning process setting the global budgets might include some participants to serve as consumer representatives. Although these participants might reflect public opinion in the planning process, they could not accurately represent consumer preferences, as these would be revealed by consumer market behavior. For planners to do this, some way would have to be found to simulate the market process—that is, to determine the expenditure cap by aggregating the cost consequences of a myriad of consumer choices, each of which would have weighed the costs and benefits of particular health services. No one has proposed a planning process to set total health spending caps that would attempt to do this (10).

According to the standard critique, the current medical marketplace is itself not well designed to reflect consumer preferences either. Because of third-party financing of the demand for care, for most choices to utilize health care consumers confront prices that fall far short of measuring the value of the resources used to produce the services. Such prices deny consumers information about the economic cost consequences of their choices to use health care. Hence, their choices do not accurately reflect their preferences about what are economically sound trade-offs between health services and other goods and services.

The Two-Person Problem

As consumers of health care, currently or prospectively, we have two roles. In our patient role, we are experiencing illness, we are seeking care, and we are using care. In our subscriber role, however, we are well and know that at some unknown time in the future we will experience illness of some unknown kind. As subscribers, we make payments to health insurance funds to provide coverage for these future events: the more the coverage, the higher the payments. Our preferences about what are appropriate rates of resource commitment to health care are very different in the two roles. As patients, we are prone to want and demand more services than we are willing to pay for as subscribers paying taxes and premiums.

(*Continued on page 105*)

(*Continued from page 17*)

This is the two-person problem. It poses a fundamental difficulty for reflecting relevant consumer preferences in the rates of resource allocation for health. Efficient markets map relevant consumer preferences into resource allocation decisions. But which are the relevant preferences: those of the patient or those of the subscriber? Which role should dominate? Appropriate management of this problem is centrally important for the design of health financing systems.

Different health financing systems manage this two-person problem in different ways. Cost containment enthusiasts favor the subscriber role simply on the ground that this will be more effective in containing costs. To the extent that planning processes setting global budgets could be designed to reflect consumer preferences, they favor domination of the two-person interaction by the subscriber role. The current medical marketplace, on the other hand, favors domination by the patient role. With prevailing patterns of third-party financing, neither patients nor providers take much account of costs of services at the point where the choice of using the service must be made.

In my view, the patient role should dominate the two-person interaction. This arrangement minimizes the regret we might otherwise experience if, as subscribers, we made decisions that turned out to be errors because they imposed unacceptable, binding constraints on our access to services as patients. In this domain where risk aversion is very important, the maximin criterion (that is, the choice of the course of action with the best of the worst outcomes) suggests that it is appropriate that the patient role should dominate. With this strategy, the worst outcome is that we end up with higher premium or tax payments than we are happy with. If, on the other hand, the subscriber role dominates, there is a more ominous worst outcome: We may find ourselves in our patient role denied medically beneficial services that we desperately want and need. In representing consumer preferences in resource allocation decisions, both health financing systems—planners setting global budgets and the current medical marketplace—have problems. The latter is to be preferred, however, on the grounds of more appropriate management of the two-person problem. This consideration, in my view, blunts the standard critique of the latter.

Implementation of Global Budgets

Where expenditure caps are proposed as a cost containment strategy, the intention is that total health expenditures, and hence the availability of services to consumers, will be less than they otherwise would have been. What is to be the role of the physician in administering the non–price rationing that necessarily must accompany the imposition of effective caps of this kind? An answer to this question is central to the evaluation of some of the difficulties to be anticipated in the administration of caps.

Many physicians regard the role of explicit, non–price rationing of medical care to be in direct conflict with their stewardship for the welfare of their patients (*11*). V. R. Fuchs, remarking that the commitment of the individual physician to the individual patient is one of the most valuable features of U.S. medical care, notes that only if such rationing is "implicit" will it be tolerable (*12*). Can the physician's role in the administration of expenditure caps then entail non–price rationing of a sufficiently implicit kind to be tolerable? One cannot be sanguine on this score. Individual physicians in their practice settings will not have responsibility for setting caps; they will inherit these from an external decision-making process. But decisions will still have to be made in the practice setting about which patients have access to those services that the externally imposed caps have assured will be in short supply. It is hard to see how doctors can avoid this role or make it into non–price rationing of an acceptably implicit kind.

This rationing role has certainly not been avoided or transmuted into implicit rationing in countries such as the United Kingdom and Canada, where health cost containment is accomplished by expenditure caps. In their observations on the health system in the United Kingdom, H. Aaron and W. B. Schwartz note: "The denial of useful or even life-saving care is hard on both providers and patients. In Britain, primary care physicians, who are forced to act as gatekeepers for the system, bear this unpleasant responsibility" (*13*). Accounts of experience under the Canadian health services system point out that physicians there likewise bear this kind of unpleasant responsibility (*14*).

A yet greater problem for the administration of expenditure caps in the United States would be the response of patients to the kind of rationing necessarily entailed. These problems now arise in other countries where cost containment relies on caps. For example, Aaron and Schwartz commented on the experience in the United Kingdom (*13*, p. 421):

> The professional and managerial classes in Britain are less willing to accept "no" for an answer than are other social classes. . . . aggressive or influential patients can often secure referrals from general practitioners for a second opinion at specialized centers or by going directly to emergency rooms for services that local doctors deem "unsuitable." As a result, per capita expenditures by the National Health Services were reported to be 41 percent higher for members of the upper two socioeconomic groups (professionals, employers, and managers) than for members of the "lowest" two classes.

The class structure of the United States is very different from that of Britain or other countries in Western Europe. The U.S. social order is relatively classless in the traditional sense of this term, and this feature of the social order is reflected in a diffuse array of U.S. social, political, and economic institutions. As a part of this, medical malpractice institutions have been developed to a far greater extent in the United States than elsewhere. These institutions confer on us in our "humble" role as patients rights that we may assert against providers in the professional class. It is not just that a denial of services by explicit rationing in the administration of an expenditure cap might result in a medical malpractice claim—although it might. The more important point is that the provider-patient relationship reflected in these institutions is such as to make the patient in the United States apt to behave as Aaron and Schwartz have characterized the upper socioeconomic groups in Britain—namely, less willing to accept "no" for an answer. If this is indeed so, caps are going to be difficult to administer in the United States.

The considerations discussed here lead to the conclusion that such caps would be, on these grounds, ill-advised in the United States. The increasingly numerous and visible proponents of expenditure caps as the best cost containment strategy have a duty to address these problems.

REFERENCES AND NOTES

1. Quoted in M. Holoweiko, *Med. Econ.* **69**, 64 (1992), *ibid.*, p. 75.
2. J. Somerville, *Am. Med. News* (11 May 1992), p. 39.
3. *ibid.* (7 April 1989), p. 37.
4. R. Saltman, *J. Am. Med. Assoc.* **268**, 774 (1992).
5. S. Jenks and G. Schieber, *Health Care Finan. Rev.*, 1 (1991).
6. S. Sonnefeld *et al.*, *ibid.* **13**, 7 (1991).
7. A. Vall-Spinosa, *Am. J. Pub. Health* **81**, 1569 (1991).
8. J. M. Weiner, *Brookings Rev.* **10**, 26 (1992).
9. U. Reinhardt, *Health Care Systems in Transition* (Organization for Economic Cooperation and Development, Social Policy Studies No. 7, Paris, 1990), p. 107.
10. D. Eddy, *J. Am. Med. Assoc.* **264**, 1737 (1990); *ibid.* **265**, 105 (1991). Eddy does not literally propose this, but he discusses a methodology that in my view might be put to this use.
11. N. G. Levinsky, *N. Engl. J. Med.* **311**, 1537 (1984).
12. V. R. Fuchs, *ibid.*, p. 1572.
13. H. Aaron and W. B. Schwartz, *Science* **247**, 418 (1990).
14. *New York Times*, 30 April 1991, p. A8.