Proposed Changes in the Organization of Health-Care Delivery: An Overview and Critique

UWE E. REINHARDT

Students of the American health-care delivery system are generally agreed that, under its present organization, the system will be unable to accommodate any of the more ambitious national health-insurance schemes now before Congress. It is argued that the current system is actually a fragmented "nonsystem" that fails to deliver the right mix of care to the right people and at the right time. As a result, it is argued, the health-maintenance services received by the average American tend to be of dubious overall quality and are unnecessarily costly. To eliminate these shortcomings, a great number of reforms have been proposed, the bulk of which, however, fall into one of the following major types: (a) a shift away from the fee-for-service mode of paying for health services and toward prepayment of comprehensive health care, (b) the substitution of paramedical for medical manpower and of capital for all types of manpower, (c) the consolidation of small provider facilities (especially solo medical practices) into larger production units, and (d) the integration of provider facilities in centrally directed regional systems.

In this essay, the various reform proposals that have been proposed at one time or another are explored against the backdrop of pertinent empirical research available at this time. This exploration leads to the disappointing conclusion that far too many of the proposed reorganization schemes—particularly the much touted idea of a nationwide network of presumably competitive Health Maintenance Organizations—appear to have been proffered more on the basis of intuition or faith than on the basis of convincing empirical evidence. At the risk of appearing timid and of exasperating the impatient reformer, the author concludes that a great deal more empirical information needs to be gathered on the behavior of the participants in the health-care sector and on the technical constraints under which that sector operates before one can confidently develop and follow a coherent blueprint for a reorganization of the American health-care system.

Introduction

Almost any of the national health-insurance proposals now before Congress, if put into operation, would add to the existing demand for health services in this country, and would therefore place an added burden on the nation's already strained health-care provider system. The nation could, of course, respond to any increase in de-

mand simply by expanding the existing system, a system which some observers regard as quite satisfactory (cf. American Medical Association, 1964, and Myers, 1970). That approach, however, is widely criticized. In the words of Harvard economist John T. Dunlop (1965:1326):

> Nothing could be worse in our society today than to say we need another three to five billion for medical care and then simply to duplicate or multiply the [provider] arrangements we now have . . . . The permanent problem is the need for more productivity . . . brought about by structural changes in the practice and organization of medicine.

Dunlop's sentiments in this respect are echoed in the writings of numerous health experts. (See for example, Ginzberg, 1969; Anne Somers, 1971; Fein, 1967; Rutstein, 1966; U.S. Department of Health, Education and Welfare, 1967; and U.S. National Advisory Commission on Health Manpower, 1967.) They are reflected also in some of the health-care proposals currently under consideration by the Congress. For example, both the Administration's and Senator Kennedy's proposals contain specific provisions aimed at a restructuring of the nation's health-care delivery system.

While there appears to be wide agreement on the need to move away from the present health-care delivery system in the United States, opinions still differ considerably on the *direction* any future reform should take. In part this lack of agreement simply reflects deep-seated ideological differences and conflicting views on what constitutes "quality" in health-care delivery. But there is also a dearth of information on the socioeconomic and medical characteristics of alternative provider systems, existing or imagined. If the choice of a new provider system had to be made at this time, it would have to be made essentially on the basis of conjecture.

This essay will in the main be a systematic review of various proposed reforms of the existing United States health-care sector. An attempt will be made to compare alternative proposals in terms of their economic attributes, to the extent that this is possible on the basis of prior empirical research or on the basis of deductive logic. However, no attempt will be made to rank alternative proposals in order of preference, or to suggest a "best" provider structure. As will become clear in the course of this essay, such a ranking can emerge only from a political consensus.

Although the essay is addressed primarily to the non-economist, it is well to keep in mind that it is written by an economist and therefore reflects an economist's view of the world. It is part of the economist's credo that human beings—including physicians—tend to be fairly responsive to financial incentives, a supposition which, in the context of a professional services industry, may strike some readers as irreverent if not offensive. To abandon that precept, however, would be to rob economic analysis of the main contribution it can make to the solution of social problems. Besides, there is ample empirical evidence that the economists' instincts in this respect are basically sound.

An Overview of the U.S. Health-Care Delivery System

The Current Organization of Health-Care Delivery

The existing health-care provider system in the United States is a conglomerate of literally hundreds of thousands of more or less independent public and private providers of care. Among the public providers are the armed forces, Veterans Administration, the Public Health Service, and certain state and local health-care facilities for the treatment of particular medical problems. The non-governmental portion of the system consists in the main of voluntary non-profit hospitals (operating about 36 per cent of all hospital beds in the United States), proprietary for-profit hospitals (operating about three per cent of all beds), privately organized laboratories and clinics, and last, but certainly not least, the 250,000 or so non-federal physicians involved in direct patient care. Of the latter, roughly 75 per cent are in office-based practice; the remainder includes residents, interns, and full-time medical staff in hospitals.

For the most part, physicians in office-based practices function as private entrepreneurs who sell their services on a fee-for-service basis. In 1969, about 20 per cent of active non-federal physicians (excluding interns and residents) were members of group medical practices of three or more physicians, although two-thirds of these groups did not exceed the three-to-four-man category. Some medical groups do render their services against an annual prepaid capitation fee, although less than 20 per cent of all group practices deliver more than one half of their services on that basis. Fee-for-service solo medical practice or two-man partnerships therefore remain the

predominant organizational form of private medical practice in this country.

With the exception of certain specialized hospitals—e.g., those for the treatment of tuberculosis or psychiatric disorders—hospitals typically base their charges for services rendered on a full-cost recovery basis, although the determination of the full costs of particular services is at best an imprecise matter. Hospitals levy their charges either on the patient directly or, more commonly, on third parties including insurance companies or government agencies paying on behalf of the patient. In contrast, the bulk of the physician's charges are still paid by patients directly with the exception of charges for surgical services that tend to be at least partially covered by some form of insurance.

In essence, then, the bulk of all medical services produced in this country are delivered to patients on a fee- or charge-per-service basis, whether or not the consumer himself pays on the same basis or indirectly through insurance premiums and/or taxes. This arrangement may strike the American as a logical choice of payment mode, as it parallels the financial arrangement accompanying the delivery of most other types of consumer services in the country. As will be indicated further on, however, in the health field, as in other service fields, the fee-for-service mode can produce some rather undesirable side effects.

Although it would be incorrect to assert—as is sometimes done—that the health-care sector in the United States is tightly *controlled* by the medical profession, it is nevertheless the case that physicians as a group have exercised a pervasive influence over the development of the sector and over resource allocation within that sector. For one, current licensure laws in the health field virtually define the physician as the consumer's only legitimate, primary contact with the health-care system. After the initial contact has been made, it is largely the physician who determines the consumer's demand for particular health services (the physician's own services, and those of hospitals, laboratories, and the pharmaceutical industry). Furthermore, through professional control over hospital accreditation and as members of individual hospital boards, physicians can, in practice, exert a powerful influence also over resource allocation within the hospital. Finally, through their individual location decisions—decisions which physicians make as private citizens—the profession collectively does determine to a large extent the geo-

graphic distribution of medical resources. Since current licensure laws effectively preclude the use of paramedical personnel as primary contacts for patients in underdoctored regions—mostly rural areas or urban slums—the location decisions of physicians indirectly define also geographic differences in the degree of access to adequate medical care.

In view of the enormous number of seemingly independent health-care providers in the United States, it may be tempting to view the health-care sector as an example of a perfectly competitive market in the classic sense. It is, in fact, a temptation to which the American Medical Association (AMA) succumbs in its well-known *Report of the Commission on the Cost of Medical Care* (1964; see, for example, pages 11, 20–21, and 32). Actually, the very nature of the relationship between the physician and his patient tends to preclude a proper functioning of the proverbial Invisible Hand in the market for medical care, for, in the conduct of his practice, the physician acts at once as the consumer's agent in defining the latter's demand for medical services *and* as the producer who meets a good part of that demand. In assuming this dominant role, the physician does not, of course, usurp the consumer's sovereignty; he simply takes over where the consumer's technical competence ends.

This relationship between physician and patient—analogous in at least some respects to the relationship of trust that is thought to exist between, say, an automobile mechanic and his technically uninformed customer—is clearly a delicate one, especially if the physician offers his services on a fee-for-service basis and as such is subject to a potential conflict of economic interests, if only in appearance. While it is widely taken for granted that the physician's professional ethics preclude the intrusion of financial incentives on his practice of medicine (for a conflicting view, see Monsma, 1970), the peculiar economic arrangement under which private practitioners in this country operate—i.e., the lack of autonomy between the supply of and the demand for medical services—should nevertheless make one reluctant to describe the market for health services as a model of competitive private enterprise.

The Alleged Merits and Demerits of the Existing System

It is generally agreed that, by virtue of a stringent selection process and of a lengthy and resource-intensive training, the average Amer-

ican physician ranks among the technically most competent medical practitioners anywhere in the world. Similarly, it is rarely disputed that such individual services as are produced and delivered under the existing health system in this country tend to be of a high intrinsic quality, even if the *mix* of these services—e.g., the mix of preventive and therapeutic care, the mix of ambulatory and inpatient care, and the mix of surgical and non-surgical intervention—is not always held to be ideal (cf. Ginzberg, 1969:98).

In addition, a society that places great value on individual freedom must surely give the American health-care system high marks for the freedom of choice it affords both the providers and the majority of the consumers of medical care. From the physician's point of view, in particular, the American system is highly attractive, for it leaves entirely up to him the choice of a specialty, the choice of the mode and location of his practice, the determination of the length of his work week, of his patient load, of the level of his professional fees, and of the style and pace with which his practice is conducted. Finally, it is generally viewed as appropriate that, other things being equal, the financial rewards earned by health-care providers tend to vary positively with their own efforts. Indeed, some observers regard this relationship between effort and reward—as embodied in the fee-for-service system—as the *sine qua non* of high-quality care, perhaps not fully realizing that so strong an assertion betrays a rather cynical view of the health-care provider's professional integrity.

On the other side of the ledger, the present system has engendered a number of problems which are sometimes said to outweigh the positive aspects enumerated above. The misgivings that have, at one time or another, been voiced in this respect can be distilled into the following summary:

1. It is held that the existing provider system, composed as it is of several hundred thousand more or less independent decision-makers ("firms"), is really a "nonsystem" that lacks effective planning and coordination. This lack of coordination is said to have resulted in

   —widespread duplication of costly facilities, equipment, and patient record systems, and hence unnecessarily high costs of health care;

—a maldistribution of medical resources, with a relative abundance of facilities in affluent urban areas and a corresponding lack of facilities in the poorer urban or rural areas; and

—lack of comprehensiveness and continuity of care.

2. It is held that current legal restrictions on medical practice, and in particular current licensure laws covering both medical and paramedical manpower, tend to

—discourage experimentation with the use of paramedical personnel (physician assistants, nurses, medical technicians) for tasks now requiring scarce and expensive physician time; and

—discourage entry of labor into the health-care sector, since current licensure laws effectively rule out the prospect of upward mobility in the health-manpower hierarchy.

While it is conceded that licensure laws do protect the consumer from unqualified personnel, it is felt that this benefit is not sufficient to offset their stifling effect on cost-reducing innovations in the organization of medical-care delivery.

3. It is suggested that the financial arrangements currently accompanying the delivery of health services, i.e., the emphasis on payment (fee or charge) -per-service, combined with the fact that the consumers are typically more fully insured for inpatient than for outpatient care tends to

—deter consumers not covered by third-party payment from seeking relatively inexpensive preventive care in the early stages of a medical condition, thus necessitating more expensive therapeutic care later on;

—encourage consumers (and their physicians) to substitute costly (but insured) hospital services for less costly (but uninsured) ambulatory care;

—bar some consumers (the lower middle class) from access to needed care altogether; and

—encourage providers to overprescribe or oversupply health services to those consumers who can afford to purchase medical care.

4. With respect to the hospital sector in particular, it is held that the prevalent full-cost reimbursement formula, combined with the fact that a hospital's prestige increases to some extent with the complexity of cases it can handle, tends to encourage the acquisition of costly facilities and equipment that are not fully used.

In addition, it is sometimes argued that the pervasive control by the medical profession over almost all facets of the health-care delivery process—e.g., the physician's voice in the management of hospitals —has tended to make the health system more responsive to the intellectual interests of the profession than to the medical needs of consumers. This allegation is sometimes accompanied by the argument that, by virtue of their training, physicians are ill-equipped to manage properly so complex a system as the United States health-care sector and that a restructured delivery system should provide for expanded lay control over the allocation of medical resources.

Finally, the entire discussion on the restructuring of our health-care delivery system is premised on the assumption that the nation will sooner or later adopt a comprehensive health-insurance plan that will effectively protect individual consumers from the financial risks currently associated with major illness. Since the present essay is concerned only with the organization of health-care production and delivery, we shall take it for granted that comprehensive insurance coverage of some form will, in fact, be introduced.

It will have been noticed that the criticisms enumerated above tend to fall into two major categories: those concerned primarily with the *quality* of the health care received by the American people as a whole, and those concerned with the *efficiency* (or costs) with which that care is being produced. In other words, almost all proposals for a restructuring of the American health-care delivery system seek to accomplish improvements along either or both of these dimensions. Since the authors of such proposals do not always make it sufficiently clear what they mean by the terms "quality" and "efficiency," it may be well to define them carefully here before

proceeding with an examination of the proposals themselves. In developing these definitions, it will also become obvious why it is quite impossible to select a "best" health-care provider system in the absence of a political consensus of what "best" means in this context.

The Concepts of "Quality" and "Efficiency" in the Context of Health-Care Delivery

The "Quality" of a Nation's Health Services

Given the present and prospective resources of the American health-care sector, any meaningful definition of "quality care" involves an implicit trade-off between two quite distinct concepts of quality, i.e.,

1. The quality of the services delivered to those to whom such services are actually available (hereafter referred to as micro-quality);
2. The effectiveness of the health-care sector as a whole in maintaining or improving the health status of the United States population as a whole (hereafter referred to as macro-quality).

This trade-off may be conceptualized with the aid of a curve such as line QN in Fig. 1.

The vertical axis in Fig. 1 is thought to represent a hypothetical, measurable index of micro-quality, observed at the nexus between individual health-care providers and their patients. The word "hypothetical" is appropriate in this context, for at this time of writing there does not yet exist a consensus among medical experts on an operationally meaningful definition of such a quality index.[1] Even so, for the sake of illustration it is assumed here that

[1]Much has been written about the problem of measuring the quality of medical services and considerable research effort is currently being devoted to that problem. For a thoughtful review of this issue, see Donabedian (1966); for specific proposals and on the definition of quality, see Morehead (1967), Peterson (1963), and Shapiro (1967). The ideal would be to base the quality index on the end result from medical treatment, although the measurement problems inherent in that approach are almost insurmounta-

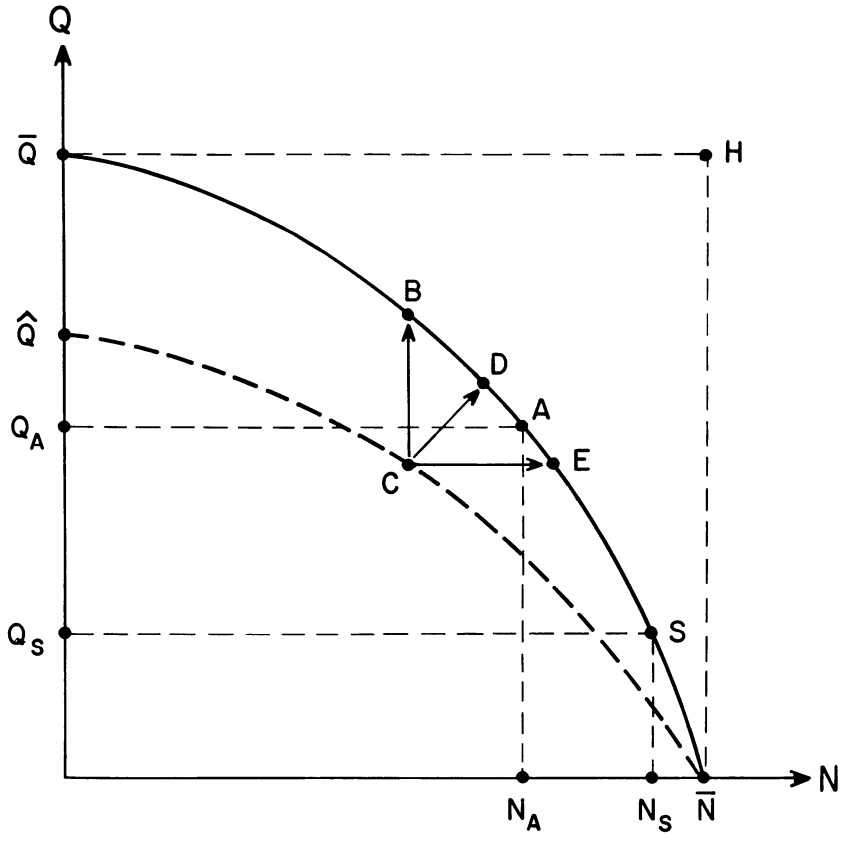

FIGURE 1. Hypothetical trade-off between micro- and macro-quality.

the development of such an index is, in fact, feasible. Furthermore, it is assumed that $\bar{Q}$ is the highest (technically attainable) micro-quality that could be offered to an individual consumer, even if all the nation's health-care resources (efficiently used) were allocated to him alone.

The horizontal axis in Fig. 1 denotes the number of persons receiving health care at all, with $\bar{N}$ representing the target population (e.g., all residents in the United States). For the sake of illus-

ble. The alternative—often used in practice—is to measure quality either in terms of the *process* of medical treatment or in terms of the quality of the inputs used in the production of medical care. Under either of these two indices, however, a quality-conscious provider is likely to confuse quality with resource-intensiveness of treatment, a fact that may result in unnecessary inflation in the cost of medical care.

tration it is assumed that those persons who do receive care all receive the same amount of care and that the latter is of uniform micro-quality.

The depicted trade-off $\overline{QN}$ is based on the assumption that medical resources are always used *efficiently*. This means, for example, that for a given micro-quality level Q_s, N_s is the maximum feasible population that can be served with the available resources or, alternatively, that Q_s is the maximum feasible micro-quality that can be attained if N_s persons are to be served with the available resources. By contrast points below $\overline{QN}$—e.g., point C—represent *inefficient* provider systems since, with a given resource base, micro-quality, Q, can be increased without reducing the number of persons served, or the number of persons served could be increased without a reduction in Q.

A good part of the criticism directed at the existing health-care sector in the United States is that, traditionally, American providers —in particular the medical profession itself—have emphasized the quality of services delivered (micro-quality) at the expense of the accessibility of health care to all (macro-quality). This allegation is inferred from the fact that access to physicians and medical facilities in this country is distributed highly unevenly both geographically and across socioeconomic classes within society. Further support for the allegation is drawn from the fact that the introduction of Medicare unleashed a surprisingly large, hitherto unmet demand for medical care.

By contrast, national health policy in Great Britain and in the Soviet Union has traditionally emphasized coverage (macro-quality), apparently at the expense of the quality of the services delivered (micro-quality). Health care in these countries is generally available to all members of society, although the quality of the care being rendered is said to fall short of American standards (Fry, 1970).

Fig. 1 illustrates a number of important points. First, it suggests the logical impossibility of the much-mouthed goal that *"all* Americans should have the *highest* quality care." This slogan presupposes a health-care system characterized by point H in the diagram. To reach that point, the nation would have to increase vastly its resource allocation to the health sector. But in so doing one would increase the highest attainable level of micro-quality (on the

vertical axis) so that pursuit of the slogan amounts to chasing a will-o'-the-wisp.

Second, Fig. 1 indicates clearly why it is *impossible* to make *objective* statements about the "overall quality" of alternative provider systems (e.g., the systems represented by points A and S in the diagram). For systems that operate subject to a given resource constraint (e.g., the constraint that not more than X% of GNP should be devoted to health care) any overall index of quality must be a weighted sum of micro- and macro-quality, and these weights can be established only through *political* consensus. Fig. 1 thus illustrates why it is possible for different experts to travel in Britain or in the Soviet Union, some returning with glowing reports about the quality of these nations' health systems and others decrying the lack of quality in British or Soviet health care. Such statements simply reflect differences in the rate at which their authors are willing to trade off micro- for macro-quality.

The gist of the foregoing discussion is that any attempt to characterize alternative health-care delivery systems in terms of their "quality" implicitly involves a rather specific view of the goals of a nation's health-care system. Since the ultimate purpose of all health-care consumption is the production of "better health," it would seem logical, first of all, to assess the quality of the services that are delivered (micro-quality) in terms of their "end result"[2] rather than in terms of either *process* or *inputs*. And if the professed goal of the nation's health system is to provide adequate care to *all* citizens, then any overall index of quality inevitably involves a trade-off between the quality of services delivered (index Q in Fig. 1) and the number of persons covered by the system (N in Fig. 1). Needless to say, an agreed-upon overall quality index of this sort does not exist at this time; indeed, the necessity of a trade-off between the two major dimensions of quality care is not usually acknowledged with sufficient candor. The best one can hope for at this point, therefore, is that speakers or authors making reference to the "quality of alternative systems" also make explicit precisely what they mean by that term.

"Efficiency" in the Context of Health-Care Delivery

An *efficient* operation or system may be defined as one which ac-

[2]In defining "end result," consumer satisfaction in general should, of course, be given some weight as well. In this connection, see also the preceding footnote.

complishes a stated objective at a minimum of resource costs. It follows that, just like the term "quality," the term "efficiency" has meaning only if it is used in connection with a clearly enunciated goal.

There are at least two distinct concepts of efficiency in terms of which one may rank alternative provider systems. First, one may rank alternative systems in terms of the total per-capita costs they generate in order to maintain each member of a target population at a minimum, adequate *health level.*[3] Second, one may rank alternative systems in terms of the costs that are incurred under them to provide a specific bundle of *medical services,* regardless of the "end result" that is achieved with these services. It is clear that these two indices will not necessarily result in the same ranking.

Just as the "end result" index of micro-quality, the health-related index of systems efficiency is, of course, much the preferred measure. Unfortunately, the index becomes operational only if the relationship between medical-care consumption and health status is clearly understood. Such an understanding, for example, would enable one to speak with confidence of "over- or underutilization" of services (e.g., surgery) or of "proper and improper" mixes of services. As was already mentioned, however, not even the medical profession itself is capable (or willing) to make such statements with any degree of certainty. At this time any health-related index of efficiency is therefore likely to contain strong subjective elements and will hence remain problematic.

The much narrower service-related concept of efficiency has the virtue of being measurable in practice, at least in a rough and ready fashion. (There is, of course, always the problem of accurate cost-measurement and of devising standard, operational definitions of individual services.) The service-related index of efficiency has the obvious drawback of taking for granted that the bundle of services to which it is related is itself an efficiently configured input into the production of health. The concept therefore eclipses from view the possibility of making economically advantageous trade-offs among different types of medical services as, for example, the substitution of early preventive for subsequent therapeutic care or of ambulatory for inpatient care. In spite of these shortcomings, how-

[3]It can easily be demonstrated that, on this definition, any provider system falling onto the trade-off curve $\overline{QN}$ in Fig. 1 must be *efficient* in this sense, as long as the micro-quality index is measured in terms of "outcome."

ever, the service-related index has been widely used in comparisons between group and solo medical practices (Newhouse, 1973) and in studies of hospital costs (Feldstein, 1968, and Ingbar and Taylor, 1968).

To sum up at this point: although almost any proposal for a restructuring of the health-care delivery system in this country originates in a desire to improve the efficiency of the system or the quality of the care provided, both the terms "efficiency" and "quality" in this context are rather elusive concepts that lack a precise and consistent interpretation. Any attempt to evaluate alternative provider structures in terms of either of these dimensions is therefore fraught with danger.

Broadly speaking, proposals for a restructuring of the United States health-care system typically envisage improvements in the efficiency in the production of better *health* (or in health maintenance). It is hoped, first of all, that the proposed changes will enhance the efficiency with which the current mix of health services —i.e., the mix of preventive and therapeutic care, of ambulatory and inpatient care, of surgical and non-surgical care, and so on—can be produced. This is efficiency in the narrow, service-related sense. In addition, however, it is hoped that an appropriate realignment of financial incentives will also lead to certain desirable trade-offs between types of care, i.e., that given levels of health status can be maintained with less costly bundles of medical services. Superimposed on any gain in service-related efficiency, it is felt, such trade-offs will assure one of substantial gains in the overall efficiency (and hence a reduction in resource costs) with which the health-care system produces or maintains given levels of health status.

In terms of Fig. 1, efficiency gains in this wider sense of the term can be represented by upward shifts in the quality trade-off curve. For example, if the system currently operates at the point C on a trade-off line characterized by inefficiency in the production of health (the broken line $\hat{Q}N$)—as is allegedly the case with the existing American system—then a move toward greater efficiency might shift the trade-off from $\hat{Q}\overline{N}$ to the new position $\overline{QN}$. It is seen that, *without any added resource use,* this move would permit the nation either to enjoy added micro-quality without a sacrifice in coverage (point B), or added coverage without a sacrifice in micro-quality (point E) or both (point D). The broad consensus

among the critics of the American health system would probably be that the system should be moved to the right, i.e., that primary emphasis should be given to added coverage. Many of these critics would probably be content even to move *downward* to the right along an unchanged quality trade-off curve should it be impossible to shift the latter upward through gains in systems efficiency. Such a dilution of the micro-quality of services would, of course, be opposed by consumers now enjoying access to adequate care of high micro-quality and probably also by the medical profession.

Admittedly the preceding exposition oversimplified matters considerably, as any abstract analysis must. Nevertheless, the diagram does serve to illustrate the relationship between changes in efficiency in health-care production and in the quality of the services rendered the nation, and we can now turn to an examination of the proposals aimed at bringing about such changes. In proceeding with that examination, it will be analytically convenient to think of the health-care system as the composite of two distinct parts:

1. The *technology* of health-care production (by which is meant the organization of medical resources—inputs—within provider facilities and the distribution of productive activity among provider facilities); and

2. The *financial arrangements* accompanying the production and delivery of care (i.e., the set of financial incentives faced by the providers and consumers of health care).

The two components are, of course, intricately related to one another and *jointly* determine the efficiency with which the system as a whole operates. For our purposes, however, it is more useful to treat them as separate parts and we shall do so in the two sections which follow. The dichotomy guards one, for example, against the not uncommon error of confusing cost savings attributable to *economies of scale* in health-care production with cost savings attributable strictly to *financial incentives.* As Pauly (1970) and others (Klarman, 1968; Bailey, 1970a; Berry, 1970) have noted, this crucial distinction is not always made in discussions on group medical practice. As a result, the effects of prepayments are occasionally taken for genuine scale economies.

Proposed Changes in the Technology of Health-Care Production

Proposals for a reorganization of the technology of health-care production generally seek to bring about one or a combination of the following three changes:

1. The consolidation of small, independent provider facilities into larger units, with the aim of reaping potential *economies of scale* in health-care production.
2. The *substitution* of relatively more abundant and/or less costly productive factors for relatively scarcer and/or more costly inputs.
3. *Increased division of labor and specialization of functions* among some types of facilities (especially in the hospital sector) and the integration of specialized units into a coordinated, comprehensive, and efficient delivery system through either full-fledged regional planning or at least the more centralized control inherent in the HMO concept.

In this section we shall be concerned only with items (1) and (2), reserving some comments on (3) for the final part of the essay.

Economies of Scale

Ever since Adam Smith's celebrated treatise on the economics of pin manufacture, economists have been fascinated by the concept of economies of scale. Precisely defined, economies of scale are said to exist when the unit cost of producing a commodity decreases as more units of that commodity are produced per period. This decrease in average per-unit costs can have at least three origins. First, certain inputs into the production process may be indivisible in the sense that they can be used to capacity only for large production runs. Second, large production runs facilitate the division and specialization of labor (and capital), a process which is generally thought to increase the efficiency with which specific tasks are performed. Third, large production runs permit bulk purchasing of certain inputs and hence offer economies in their procurement.

The notion that, up to a point at least, economies of scale exist in almost all conceivable production processes is certainly plausible.

It is therefore not surprising that the phenomenon is often attributed to health-care production as well. In applying the concept to that setting, however, one runs into at least two major problems. First, it is extremely difficult in most cases to define the unit of medical service in terms of which the suspected economies can be measured. Empirical research on the extent of economies of scale in health-care delivery is thus severely handicapped. Second, consumers of health care are not totally indifferent to the *manner* in which the care they consume is produced. While consumers normally do not care whether a pin is hand-crafted or produced on an assembly line, they generally do have strong misgivings about so-called assembly-line medicine. If consumer satisfaction—as distinct from the objective, medical merit of treatment (e.g., an injection)—is to be considered in evaluating the quality of care, then a comparison of costs at different rates of output of observable units of service is not particularly helpful for the formulation of policy.

These methodological difficulties notwithstanding, a considerable amount of economic research has, in fact, been devoted to identifying the presence or absence of economies of scale in health-care production. The bulk of this research has focussed on the hospital sector. (See Feldstein, 1968; Ingbar and Taylor, 1968; Cohen, 1967; Lave, 1966; Mann and Yett, 1968; Ro, 1968.) The most frequently used index of scale in that context has been not some rate of output—the dimension in terms of which economies of scale really ought to be defined—but rather the number of beds per hospital.

Most empirical investigations of hospitals have led their authors to the conclusion that the production of hospital care is, indeed, characterized by economies of scale. Unfortunately, the estimated point of minimum costs per patient day—i.e., the estimated optional scale—has ranged from a low of 160 beds (Cohen, 1967) to a high of 900 beds per hospital (Feldstein, 1968). This range is clearly much too wide to be of use to policy makers.

On the other hand, some researchers have failed to discover any pronounced scale effect in hospital-care production. The Panel on Hospital Care of the U.S. National Advisory Commission on Health Manpower (1967, Vol. 2:162) for example, emerged from its survey of 12 "distinguished" hospitals with the conclusion that "no relationship can be discerned between cost per patient day and hospital size." And one pair of researchers (Ingbar and Taylor, 1968)

has estimated an *inverted* U-shaped cost curve with the maximum-cost hospital in the 150–200-bed range. Although the authors do not set great store by this result, concluding only that average hospital cost curves do not decrease monotonically in the 30–330-bed range, their result nevertheless lends credence to the agnostic conclusion reached by one reviewer of hospital-cost studies (Berki, 1972:115):

> In summary, then, the question can be posed: What are the shapes of the short-run and long-run cost functions for hospitals? Are there economies of scale? The answer from the literature is clear: "The exact general form of the function is unimportant" but "whatever its exact shape," and depending on the methodologies and definitions used, economies of scale exist, may exist, may not exist, or do not exist, but in any case, according to theory, they ought to exist.

It would be tempting to add to Berki's conclusion the admonition that "more work needs to be done on this problem" were it not for the fact that the presence or absence of scale economies within given types of hospitals is probably not the most pressing policy issue at this time. While it is true that, if one knew the precise shape of the cost-output relationship for a given hospital, one could easily identify the optimum scale (i.e., the rate of output at which average costs per unit of output reach a minimum), it is also true that any randomly selected group of hospital administrators could probably give one as good a feel for the approximate optimum size as could be learned from a proper cost function. Even in the absence of reliable information on this function, it is known, for example, that there is widespread duplication of underutilized equipment in the hospital sector, an excess capacity that is probably most glaring in connection with facilities for open heart surgery (Lee, 1971), but may be presumed to exist also for less exotic equipment. (In this connection, see also the very interesting thesis on hospital behavior recently advanced by Martin S. Feldstein, 1971.) The really important question may therefore not be "What is the optimum size of a given type of hospital facility?" (a question that may not even be meaningful in the face of a geographically limited demand) but "What overall configuration of hospital facilities within an entire re-

gion is capable of meeting that region's demand for hospital care at minimum overall costs?" Hospital cost functions of the traditional variety can shed some light on this question, but they are not a particularly helpful tool for solving the problem.

Although some uncertainty over the existence or absence of scale effects in the hospital sector is generally acknowledged, the presence of such effects in the production of *ambulatory* care has long been taken for granted. According to the proponents of group medical practice, the case for the existence of scale effects is so strong as to require no empirical proof. At first glance this argument is persuasive. For example, it must surely be the case that, in comparison with the traditional solo practice, the group-practice setting permits a more extensive division of labor among the physicians' staff, that managerial and record-keeping functions can be delegated to business managers or to clerical personnel, and that groups can use X-ray and laboratory facilities more nearly to capacity. Furthermore, it may be the case that the group-practice mode is relatively more hospitable to technological innovations, such as the use of on-line computers for diagnostic purposes. Finally, group practices may even achieve greater economies in the procurement of drugs (McCaffree and Newman, 1968) and medical supplies than can physicians in solo practice.

While such deductive reasoning has a certain intuitive appeal, there exists at this time little empirical evidence to validate one's intuition.[4] Some proponents of the group concept (e.g., Boan, 1966) have cited the typically higher net incomes of group practitioners as prima facie evidence for the existence of scale effects. However, income differentials of this sort may simply reflect the fact that a group practice can internalize the production of (and the profit margins from) certain ancillary services that solo practitioners refer to outside facilities. A mere shift in the locus of production, however, is not by itself evidence of economies of scale. In this connection, Richard Bailey (1970b:271–272)—one of the more outspoken doubting Thomases on this issue—deserves extended quotation:

[4]In his study of self-employed American physicians, Reinhardt (1970) did find that physicians in small (2–3 man) single-specialty groups did, on average, see 5 to 6 per cent more patients than did solo practitioners working identical numbers of hours and having identical auxiliary staffs.

> Laboratory testing at a volume which can efficiently utilize automated equipment and full time technologists shows clear evidence of increasing returns [to scale]. Such equipment is justified economically only as part of the diversified product offering of a medical firm in which there are a sufficient number of physicians to generate demand internally. The important point here, however, is that the individual physician's productivity is not affected by possessing such equipment. He can usually obtain laboratory, X-ray, and other technical products from other firms which specialize in them. . . . These services are not essential components of the physician's production function; they need not be produced jointly with physician services. Rather, they are complementary goods which can be produced apart from the physician's organization—often in firms where increasing returns of scale are [*sic*] clearly recognizable.

The implication of Bailey's quite interesting observation cannot be emphasized too strongly. In effect, he suggests not only that economies of scale in the production of *physician services* are improbable, but also that the production of ancillary services on the group's premises may be a less efficient (hence more costly) procedure than the solo practitioner's reliance on the very large and highly specialized outside providers of such services. This conclusion is certainly consistent with Joseph Newhouse's (1973) recent finding that, in a sample of solo and group practitioners in Los Angeles, the group practitioners tended to have higher overhead costs per patient visits than did their colleagues in solo practice. Newhouse emerges from his research somewhat skeptical about the relative merits of group practices.[5]

It must be added here that neither Bailey nor Newhouse make much of the possibility that the quality of medical services produced by group practices is higher than that in solo practices, a suggestion that has frequently been offered by the proponents of the group-practice mode and justified by them by the more extensive peer review possible in the group setting. Furthermore, the authors do not take into account the possibility that the group-practice mode may offer the patient significant economies in the use of his time. In-

[5]Actually, Newhouse's finding is not really strong evidence against the group-practice mode, since "visits" in the group setting do tend to include more "home-produced" ancillary services than do visits in solo practice. Even so, one would have expected items such as Billing Cost per Visit or Medical Record Cost per Visit to be lower in group practices. In fact, they were higher.

deed, only rarely have economists treated the time costs to patients explicitly as an element of overall treatment costs. Since the patient's time is certainly not a free resource from society's point of view, any group-solo comparison excluding that resource cost is suspect from the outset.

This is not the proper medium for an extended review of the literature on the matter of scale economies. But enough has surely been said to indicate that there is still considerable uncertainty on (a) suitable methods of defining and measuring the phenomenon in the context of the hospital or of medical practice, and (b) the prevalence or absence of scale economies in health-care production. In connection with the delivery of ambulatory care, the problem certainly deserves more extended and more careful analysis. Whatever past research there has been on the question has so far failed to yield the hard data on which policy can be based.

Factor Substitution

Quite aside from the question of economies of scale, there is much speculation and mounting empirical evidence that health care in this country is often produced with unnecessarily expensive combinations of medical resources. The proposed remedy is a substitution of relatively more abundant and/or less costly for relatively scarcer and/or more costly productive factors. Two distinct types of factor substitution have been suggested:

1. The substitution of ambulatory care for inpatient care or, more precisely, the substitution of productive inputs used by ambulatory-care facilities for productive inputs used by hospital facilities.
2. Within both the ambulatory-care facilities and hospitals, the substitution of less trained manpower for more highly trained manpower (e.g., the substitution of paramedical for medical personnel, or the substitution of orderlies for nursing staff), and the substitution of capital equipment (including computers) for some or all types of health manpower.

The question to be addressed in this section is whether the proposed substitutions are, in fact, *technically feasible*. What particular incen-

tives are likely to induce providers to make the desired substitutions is another matter entirely. That question will be taken up in a subsequent section on financial arrangements.

The Substitution of Ambulatory for Hospital (Inpatient) Care. Hospital care has always been the most resource-intensive type of health care. It is also the most rapidly rising component of the Medical Care Price Index, with the latter itself being one of the more rapidly increasing items in the overall Consumer Price Index. It is therefore not surprising that there is much concern over the possibility that hospital use in this country is excessive. To quote the previously cited Panel on Hospital Care (U.S. National Advisory Commission on Health Manpower, 1967, Vol. 2:135):

> In many instances the services and facilities required to treat a specific medical condition are available only in the hospital. . . . In other instances, however, where there is no medical justification for hospitalization, the patient may still be placed in the hospital either *to add to the convenience of the physician or to provide the patient with insurance coverage for diagnostic tests not covered on an outpatient basis.* There is evidence that such unnecessary hospitalization is of significant extent. [Italics added.]

Just what constitutes a "proper" rate of hospital utilization is, of course, not an easy matter to determine. To be sure, it has been suggested on a number of occasions that the lower incidence of elective surgery (e.g., tonsillectomies, appendectomies, and hysterectomies) in Great Britain, where physicians are salaried, is prima facie evidence of "unnecessary" surgery in the United States and hence of "unnecessary" hospitalization. But this argument may be countered with the proposition that what is billed as "overutilization" here really reflects "underutilization" elsewhere. To quote Odin Anderson (1964:728):

> The present patterns of use of hospital care in North America and Europe make no sense, i.e., they show no association with any given set of circumstances. The obvious conclusion is that the volume of "proper" hospital care is highly elastic, so elastic that I feel that I can generalize there is no "proper" level of use of hospital care that can be established as a standard.

Anderson's professed agnosticism in this respect is tantamount to the assertion that society does, in fact, have considerable leeway in choosing the mix of inpatient and outpatient care produced to achieve given levels of health maintenance, which is another way of saying that trade-offs between the two types of care are held to be technically feasible. The assertion finds empirical support in the fact that prepaid group-practice plans such as the Kaiser Foundation Health Plan in California or the Health Insurance Plan of Greater New York (HIP) tend to report significantly lower hospital utilization rates than are observed for the United States as a whole. For example, the U.S. National Advisory Commission on Health Manpower (1967, Vol. 2:209) reported that, in 1965, the ratio of physician visits to hospital days for Kaiser members was about 8.5; the corresponding ratio for California as a whole was only 5.5. It is not unreasonable to assume that the relatively less hospital-intensive treatment given to Kaiser members reflects the economic incentives under which that plan operates: since the plan's revenue is predetermined, there is every incentive to trade off relatively more costly inpatient services for ambulatory care.

The data cited raise the more general question of whether the overall quality of care delivered by prepaid group practices is at all comparable to that delivered by the providers under the more traditional Blue Cross–Blue Shield plans. After a comprehensive survey of the evidence on this question, Donabedian (1969) concludes that the quality of care delivered by prepaid groups tends to be, if anything, higher than that delivered under the more traditional fee-for-service system. To the extent that this assertion is correct, the experience of present prepaid group practices might then be viewed as a clue to the trade-off between inpatient and outpatient care that is technically feasible and economically wise. On the other hand, at least some studies on this issue have failed to indicate the predicted difference in hospital utilization between prepaid plans and more traditional fee-for-service plans (Densen *et al.*, 1962; Foundation on Employee Health, Medical Care and Welfare, 1962). These studies have persuaded some authors (e.g., Klarman, 1969—an excellent review and critique of the pertinent literature) that the empirical record on hospital use under prepayment is far more conclusive and that the sources of observed variation in hospital use must be more fully understood than they

are now before the experience of the prepaid plans can be accepted as a model for efficient resource allocation. Among the questions that future research should seek to answer are:

1. To what extent is it actually medically sound to substitute ambulatory for inpatient care; and
2. To what extent can such substitution be viewed as true economy from society's point of view.

The latter of these questions is a particularly intriguing one. For example, it may on the surface appear to be desirable to eliminate an elective hospital stay devoted, say, to diagnostic procedures, and to render ambulatory care in its stead. It is conceivable, however, that a rather large number of separate ambulatory visits will thus be substituted involving, perhaps, some duplication of effort and certainly much inconvenience to the patient. It is well to remember here that the hospital setting also may be able to yield those economies of scale more commonly said to give group medical practices an edge over solo practices. In sparsely populated areas that cannot support several group practices side by side, heavy reliance on the hospital therefore may be a quite sensible substitute.

In this connection a comparison of Canadian and United States data may be illuminating. Tables 1 and 2 present a number of utilization and cost indices drawn from a variety of sources. It is seen that, relative to the United States, Canada makes much heavier use of hospital facilities and is hence able to make do with far fewer physicians per capita. Although the trade-off chosen by Canada strikes one's long-conditioned intuition as uneconomic, that conclusion is not supported by the cost data in Table 2. If one adjusts these cost data for the prevailing exchange rate ($1.08 Canadian for $1 U.S.), then the indicated American per capita cost of the major health components shown in Table 2 is roughly 50 per cent higher than the corresponding Canadian figure. A substantial portion of this difference is, of course, attributable to differences in the costs of medical inputs, primarily the wages, salaries, and net earnings of health personnel. A comparison of these earnings data suggests that perhaps 40 to 45 percentage points of the overall difference of 50 per cent can be so explained. The interesting point is that after these adjustments have been made, the Canadian per-cap-

ita cost figure is not *higher* than the American figure, a result that would have been expected.

TABLE 1

Comparative Statistics on the Canadian and United States Health-Care Systems, 1968

	United States	Canada	Province of Saskatchewan
Number of hospital beds per 1,000 population	8.3	10.2	11.1
Number of patient days per capita	1.4	2.0	2.2
Number of physicians per 100,000 population	163	139	119
Number of physician visits per capita (excluding hospital visits)	4.3	NA	4.4

Sources: For U.S. data, the U.S. Bureau of the Census, *Statistical Abstract of the United States, 1971:* Tables 93, 94, 96, and 100. For Canada as a whole, the Dominion Bureau of Statistics. *Canada Year Book 1970–71,* Chap. 6: Tables 3, 6, and 18. For physician visits in Saskatchewan, the Saskatchewan Medical Care Insurance Commission, *Annual Report 1969.*

TABLE 2

Per-Capita Expenditures on Health-Care Components United States and Canada, 1967

	United States (US$)	Canada (Can $)
Hospital care	$89.21	$93.01
Physician services	50.90	33.60
Dental services	16.63	9.16
Prescription drugs	27.97	11.72
Total	$190.44	$147.45

Sources: For U.S. data, the U.S. Bureau of the Census, *Statistical Abstract of the United States, 1971:* Table 85. For Canadian data, the Committee on Costs of Health Services, *Task Force Reports on the Cost of Health Services in Canada,* Vol. 1, p. 11.

It will also be noted from Table 1 that, in 1967, the number of physician visits per capita in the Canadian province of Saskatchewan was almost identical to the corresponding U.S. figure, in spite of the fact that there were far fewer physicians per capita in Saskatchewan than were available in the U.S. It would appear from these data that the Saskatchewan physician's heavy reliance on the hospital enables him to carry a much higher patient-visit load at the office than is carried by his U.S. counterpart. Although recent data on the use of paramedical assistants in physicians' offices are not available for Canada, it is known that, in 1960, Canadian physicians tended to employ fewer such aides than did their American counterparts. In short, then, the Saskatchewan physician appears to have substituted the collective services of the hospital for an expansion of his own office staff. Whether this arrangement ultimately involves higher social costs than would obtain were Saskatchewan to adopt the overall U.S. mix of inpatient and ambulatory care is as yet an open question, but an interesting one and one that merits careful analysis.

Factor Substitution among Types of Manpower and between Capital and Manpower within Hospitals and within Ambulatory-Care Facilities. Even if the often advocated substitution of ambulatory for inpatient care were desirable and were in fact made, further gains in economic efficiency could almost certainly be achieved if, *within* each of the hospital and ambulatory-care sectors, less expensive resource inputs were substituted for more expensive ones. As noted above, this would involve the substitution of less trained for more highly trained manpower and the increased support of all types of manpower by capital equipment.

Data for the United States health-care sector as a whole indicate that a substitution of less trained for more highly trained manpower has, in fact, occurred at a steady rate over the last several decades. In his study of the job structure of health manpower, for example, Jeffrey Weiss (1966:117) found that:

> If the 1950 job coefficients for health manpower had been maintained [one decade later], an additional 100 thousand health jobs with a high level of job content [high level of training] and 13 thousand health jobs with a middle level of job content, would have been required to produce the 1960 output of health services. Instead, 117 thousand jobs with a low level of job content were substituted for these 113 thousand jobs.

In connection with physician manpower, Weiss concluded that had the 1950 manpower proportions used in health-care production been maintained in 1960, there would have had to be 46,000 more active physicians in 1960 than there actually were!

While these historical data do demonstrate the technical feasibility of factor substitution in the health-care sector, the question remains how much further this substitution can be pushed in the future.

With the notable exception of work by Martin Feldstein (1968) and Karen Davis (1969) there is very little published research on production functions (i.e., the input-output relationship) for hospitals. There is some evidence that the utilization of health personnel within the United States hospital sector does vary somewhat among hospitals (Kehrer, 1970)—apparently more so among nonproprietary than among proprietary hospitals (Clarkson, 1971)[6] —but no study of the American hospital has so far identified the degree of substitutability among various hospital inputs and the economically most efficient mix of inputs.

In the final analysis, the effective constraints on any factor substitution in the hospital sector may, of course, be found not to be *technical* at all, but instead to result from certain institutional factors characteristic of that sector. Foremost among the latter is what Greenfield (1969) has referred to as the "increasing professionalization of health-manpower" in the hospital. This process has produced a rather rigid occupational hierarchy with fairly impenetraable barriers between occupations, and with specific tasks associated with each rung on the hierarchical ladder. Not only does this occupational rigidity tend to preclude any vertical mobility of health personnel, but it also renders the reallocation of tasks across occupational boundaries much more difficult than it is in most conventional productive enterprises. In the face of these institutional barriers to economic efficiency, it is perhaps not surprising that the *technology* of hospital-care production has been deemed of little empirical relevance.

In contrast to the hospital sector, the utilization of health manpower in ambulatory patient care has received considerable atten-

[6] This finding has been taken as evidence that proprietary hospitals, more exposed as they are to the discipline of the market place, tend more nearly to adhere to a market-determined optimal input combination than do nonproprietary hospitals. In this connection, see Clarkson (1971).

tion in recent years. It is widely believed that the average American physician today—whether in solo or group practice—still performs far too many tasks that could safely be delegated to paramedical or clerical assistants. For example, in their survey of pediatricians in the United States, Yankauer *et al.* (1970:36) found evidence of "substantial inefficient utilization of both trained medical and nursing manpower in pediatric office practice," and that "respondent opinion was clearly in favor of changing this unsatisfactory state of affairs by delegating many patient caretaking activities to trained allied health workers." Among other findings, the authors (Yankauer *et al.*, 1970:45) offer the following highly interesting observation on the economics of large-scale specialty groups:

> Multispecialty groups, as defined in [our] survey, are a mixed lot, and the proportion that include prepayment for services is unknown. Nevertheless, patient care task delegation by 394 pediatricians in such groups was distinctly less common than task delegation by 739 solo practitioners, 277 in two-man, 276 in three-man, and 130 in four-man pediatric specialty [single specialty] groupings . . . . Survey data show that one key to greater efficiencies lies not in the *size* or *financing* of the setting, but in the setting's employment and utilization of nonmedical health manpower for patient caretaking. [Italics added.]

This conclusion underscores, once again, the considerable uncertainty that still surrounds the question of scale economies in group practices. It also serves to shift the emphasis away from efficiency gains through scale economies to efficiency gains through factor substitution. Since the medical profession itself still exhibits a very strong preference for the solo-practice or small-partnership setting, it may be much more expedient and efficacious to focus public policy on ways to encourage greater use of paramedical aides by solo practititioners rather than on ways to encourage the consolidation of solo practitioners into large-scale groups or Health Maintenance Organizations. This proposition certainly deserves careful consideration in any attempt to restructure the health-care provider system of this country.

An idea about the potential effect of health-manpower substitution on physician productivity may be gained from recent empirical work by Reinhardt (1972a), Smith *et al.* (1972), and Pondy

(1971). In the first of these studies, an attempt was made to estimate the quantitative impact of the substitution of traditional types of aides (nurses or medical technicians) for the time of self-employed physicians in solo practices and two-to-three-man partnerships. The analysis suggested that the delegation of tasks to such personnel can have a rather profound effect on the weekly visit rate reported by physicians. On the basis of the estimated relationship between physician hours, paramedical aides, and weekly office visits, it was found that the typical physician would maximize his income if he employed roughly four aides rather than the observed actual average of less than two aides (the actual average is for the period 1965–67). Expansion of the auxiliary staff to the estimated optimum level was estimated to increase the physician's own productivity by roughly 30 per cent (see Reinhardt, 1972a:64).

In their analysis of task delegation, Smith *et al.* (1972) focused primarily on the role of newly trained "physician assistants" in private medical practice. The authors analyzed highly disaggregated utilization data and staffing patterns in six general practices within an activity-analysis framework. Their analysis suggests that a judicious delegation of tasks from the physician to traditional aides and to physician assistants is capable of increasing the physician's average productivity by about 75 per cent (Smith *et al.*, 1972:218). This estimate is corroborated by Pondy's (1971) evaluative study of the Duke Physician Assistants Program. Pondy found that the use of physician assistants tends to increase the productivity of the employing physician by between 40 and 70 per cent. These productivity estimates take on added meaning when it is recalled that a mere *four per cent* increase in the average productivity of American physicians would add more to the available supply of medical services than would the entire current graduating class from American medical colleges.

It must be emphasized here that the productivity estimates indicated are based on task delegation currently observed within the existing legal or conventional restrictions on the extent of this delegation. Some eminent students of medicine are convinced that the use of paramedical aides in medical practice could be pushed far beyond the limits now being observed. In his testimony before the U.S. Senate Health Subcommittee, for example, Professor William B. Schwartz, M.D. (1971:448), Chief of Medicine in the Tufts University School of Medicine, has argued:

> There seems little question that physician's assistants, taking patient histories, carrying out physical examinations, and administering intravenous fluids, can greatly augment the efficiency of the doctor and thus contribute to the relief of the physician shortage. I would like to suggest, however, that this approach to the use of nonphysicians is much too limited. It is conceivable that by going further we can produce a revolution in the use of health manpower in which the physician's efforts are truly reserved for those tasks which require his high level of skill, education and intellect. I am suggesting, in other words, that if we undertake a rigorous analysis of what the doctor does, we will almost certainly find that a substantial number of his tasks, now considered sacrosanct, could be done instead by skilled technicians who could be quickly trained for single specialized tasks: for example, to diagnose and treat simple fractures, remove an appendix, strip varicose veins, carry out therapeutic abortions, or perform needle biopsies of the kidney and liver. Such new uses of manpower could well free a significant additional fraction of the physician's time.

At this time of writing, the existing licensing restrictions on tasks assignable to paramedical personnel and, in particular, the locus of legal responsibility in malpractice suits, effectively preclude the implementation of Professor Schwartz's visionary proposal, and, indeed, may even explain why the average American physician has been so timid in matters of task delegation and seems to employ fewer than the optimal number of aides. Experimentation with the rather bold task delegation advocated by Schwartz therefore awaits a restructuring of the law now surrounding the practice of medicine. And widespread adoption of his concept of medical practice presupposes also a rather drastic shift in patients' attitudes toward medical care. After all, one could hardly expect the chief providers of ambulatory care, i.e., the physicians, to rationalize their practices if that course of action opened them to increased risk of malpractice suits or to a loss of patients. Just as in the hospital sector, the effective limit to factor substitution in ambulatory patient care appears now to be set by institutional rather than strictly technical factors.

It is appropriate to end this section with a few comments on the substitution of capital for health manpower in both the ambulatory-care and the hospital setting. Electronic computers are without doubt the primary type of capital equipment that comes to mind in

this connection. The most obvious use of computers is, of course, the maintenance of patient records and the storage and processing of operational data in general. In these areas the substitution of computers for human effort has already proceeded for some time. Some students of the health-care delivery process, however, see the computer as a far more revolutionary tool. To quote, once again and at length, the adventurous Professor Schwartz (1971:444–445):

> [It] seems probable that in the not too distant future the physician and the computer will engage in frequent dialog, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action. One may hope that the computer . . . will [thus] help free the physician to concentrate on the tasks that are uniquely human, such as the application of the bedside skills, the management of the emotional aspects of disease, and the exercise of good judgment in the nonquantifiable areas of clinical care.
>
> The computer, used in this manner, might also open the way to quite different means of employing nonphysician manpower. . . . Computer-supported [nonphysican] health-care specialists, aided by a variety of automated devices for history taking, blood analysis, and other procedures, and trained to perform a careful physical examination, might take over a large segment of the responsibility for the delivery of primary medical care. . . . [L]inked to regional consulting centers by appropriate display devices, the new breed of health-care specialist could make a major contribution to the resolution of the seemingly insoluble problem of maldistribution and shortage of physician manpower.

It is clear that the author proposes a health-care provider system that is fundamentally different from what it is today. And, once again, the evolution of such a system must await certain psychological and legal changes which make it acceptable to both patients and health-care providers.

Alternative Financial Arrangements in Health-Care Delivery

It was mentioned in the introduction that the financial incentives built into the existing health system in the United States tend to generate undesirable side effects on both the technology of health-

care production and on patterns of health-care utilization. Many critics of the existing system have therefore called also for drastic changes in the financial arrangement accompanying the delivery of health care. These changes have usually been incorporated into proposals for a national health insurance system. Their intent will become clearer if we survey the range of conceivable financial arrangements at a more abstract level.

A Menu of Alternative Financial Arrangements

Broadly speaking—and at the risk of some oversimplification—one can distinguish among at least six major types of financial arrangements that could be superimposed upon any particular technology of health care delivery. In Table 3 these arrangements are identified by the letters A, B, C, and so on. The term "marginal" in this table is understood to mean the "additional profit (or cost) generated when one extra unit of service is delivered to patients."

TABLE 3

Some Conceivable Financial Arrangements of Health-Care Provider Systems

PROVIDERS FACE:	CONSUMERS FACE:	
	Positive Marginal Out-of-Pocket Costs[a] per Unit of Service	Zero Marginal Out-of-Pocket Costs per Unit of Service
Positive net marginal profits per unit of service	A	B
Zero net marginal profits per unit of service	C	D
Negative net marginal profits (positive costs) per unit of service	E	F

[a] The term "out-of-pocket" costs is synonymous with "monetary" costs. It is clear that the consumption of medical services always imposes certain marginal time and psychic costs on patients.

Consumers whose use of medical services is fully covered by third-party payment come under arrangements B, D, or F. Also included in group B-D-F are insured patients whose contracts provide for a fixed deductible per period but whose expenditures have already exceeded this deductible.

Consumers who pay for all or at least part of the cost of medical services rendered them fall into categories A, C, or E. This group includes persons whose insurance contract contains a coinsurance clause, who face a fixed utilization fee per physician contact, or who have not yet reached the full deductible they must assume under their insurance contract.

Physicians who render their professional services on a fee-for-service basis, and most non-federal hospitals, operate under either arrangement A or B. It may be suggested that hospitals which are paid for their services on an average-total-cost basis really fall into either cell C or D. However, since many hospital costs are fixed in the short run, it is legitimate to treat the difference between revenues per unit of service and average *variable* costs per unit of service—i.e., the so-called contribution to overhead—as something akin to a profit margin. At any point in time, a hospital can be expected to respond to this contribution-to-overhead margin just as it would to a genuine net profit margin per unit of service.

Salaried physicians and federal hospitals come under either arrangement C or D. Finally, any provider who had contracted with subscribing consumers to furnish them with satisfactory and comprehensive health care in return for a fixed prepaid annual capitation fee falls into either category E or F. The prepaid group-practice plans now operating in this country (e.g., Kaiser and HIP) and the Health Maintenance Organizations (HMO's) envisaged in the Administration's proposed health-care program are examples of type E or F providers. Arrangement E prevails under such plans if subscribers pay a utilization fee per visit. If no utilization fee is charged, the plans operate under arrangement F.

The classification of systems in Table 3 is, of course, not exhaustive; there are a number of arrangements that fall somewhere in between the limiting types shown in the table. For example, there exist remuneration schemes under which the provider's income or profit does not vary *directly* with each unit of service delivered, but does vary *indirectly* as a function of the average volume of services

(or their quality) he has rendered in the past. Such providers would fall somewhere between cells A-B and C-D. Similarly, consumers may find their health insurance premiums to be a function of their past utilization of health services. For such consumers the out-of-pocket cost of health services may not vary *directly* with the number of services they consume, but still be determined indirectly by a lagged function of overall utilization rates. These consumers would therefore fall somewhere between cells A-C-E and B-D-F.

Financial Incentives and Utilization Patterns

According to the familiar Law of Demand, the quantity demanded of almost any good or service per unit of time tends to vary inversely with its price, *other things being equal.* Few students of the health-care sector today would argue that this law does not apply also to the broad category of services lumped together under the name "health-care," even if the demand for certain acute care (e.g., the setting of a broken leg or the response to any trauma) is totally insensitive to price. What is not known accurately is precisely how sensitive the demand for health care is to price (i.e., the price elasticity of demand).[7] Empirical research on the question has generally pointed to a rather low price sensitivity for medical care as a whole. It would be more meaningful, however, to estimate price elasticities for particular medical procedures. A priori, one suspects that the price sensitivity of elective procedures—e.g., routine visits to the physician's office and most preventive care—may be quite high. But whatever the actual case may be, it is undoubtedly safe to advance the proposition that, other things being equal, consumers in categories B, D, or F of Table 3 will tend to demand more medical services in toto than will consumers in categories A, C, or E.

A priori, one would expect that providers of services whose net profit or net contribution to overhead varies positively with the number of services they deliver (i.e., providers in cells A or B) will tend to maximize the number of services they sell per unit of time. Conversely, one would expect that providers whose income varies inversely with the services rendered per period would seek to minimize the delivery of services subject, perhaps, to certain legal constraints or to the necessity to maintain a good reputation over time.

[7] The price elasticity of demand for a commodity is defined as the ratio of the percentage change in quantity demanded to a given percentage change in price.

Simply to graft this quite general proposition on to the specific context of health-care delivery, however, is to suggest that the providers of medical services (e.g., physicians) will readily compromise their professional ethics for the sake of financial gain. That allegation has, of course, been made on a number of occasions. For the most explicit statement of this hypothesis see Monsma (1970). Other authors who have suggested a positive correlation between utilization rates and the marginal profits faced by physicians include Wolfe and Badgley (1967) and Pauly (1970). Those who argue that the marginal profits faced by health-care providers do, on balance, tend to influence the rates of utilization that are prescribed typically point to the fact that, for roughly similar populations at risk, per capita rates of surgery are usually much higher under a fee-for-service arrangement than under a system in which physicians are salaried (Monsma, 1970:150–157). Needless to say, that evidence is not universally accepted as conclusive. At this point, it is perhaps more appropriate and certainly safer to offer the less offensive proposition that, under arrangements A–B, providers of health care are less likely to discourage any overutilization of health services—or that they are more likely to apply complex and costly treatments in the management of a particular disease—than they would be under either arrangements C-D or E-F. Alternatively, it may be assumed that under arrangements E or F, providers of health care are more likely to be conservative in prescribing or applying medical services to the treatment of a medical condition than they would be under arrangements C and D or A and B.

The preceding behavioral assumptions permit one to make some broad generalizations concerning the utilization patterns likely to obtain under the various financial arrangements in Table 3. Within any row in Table 3, one would expect the utilization of health services to increase (other things being equal) as one moves to the right; within any column of the table, one would expect utilization of health services to decrease (other things being equal) as one moves from the top to the bottom row. Under arrangement E, the utilization of services is likely to be lowest, and under arrangement B it is likely to be highest, other things being equal. Under systems A and F, the economic interests of patients and providers are diametrically opposed and it is not clear, a priori, which system is likely to generate higher utilization rates. One's knowledge of the relative power of patients and providers, however, leads one to sus-

pect that in most instances utilization rates under A will tend to be higher than those under F.

Financial Incentives and Efficiency in the Production of Health Care

The preceding generalizations were predicated on the implicit assumption that the provider's choice of a production technology is not influenced by the manner in which he is paid for his services. This assumption will now be relaxed, since the manner in which providers are paid for their services may be thought to exert a strong influence by the relative efficiency with which these services are being produced. Our main concern will therefore be to distinguish among the rows of Table 3. The crucial distinction here is between rows one and two, on the one hand, and row three on the other.

It may be thought—and it has sometimes been argued (Pauly, 1970:115)—that a provider who delivers health care at positive marginal returns has every incentive to produce whatever services he delivers at minimum cost. This supposition will be recognized as a classical tenet of economic theory. Implicit in that supposition, however, is the further assumption that the discovery and maintenance of the most efficient (least cost) production mode is not associated with any special effort on the part of the entrepreneur. In the real world, the search for efficiency is not costless, and the relative degree of success in that search is undoubtedly a function of the economic pressure under which producers operate. Producers who must take prices as given and seek to maximize their income will inexorably be driven toward the most efficient configuration of productive inputs and toward the most effective use of whatever inputs are hired. On the other hand, it is reasonable to suppose that if the higher inefficiency, if any, can easily be passed on to consumers, then the incentive to produce efficiently is greatly diminished. In concrete terms this means, for example, that health-care providers who can raise their professional fees or per-diem charges almost with impunity are not likely to be as economy-minded as are providers who are constrained by fixed, market-determined or negotiated schedules of fees or charges, or who sell their services on a prepaid capitation basis.

Since in this country there have never been any great obstacles to increases in physician fees or hospital charges, the preceding

comments suggest that one would not expect providers in cells A to D of Table 3 to be greatly concerned over the efficiency (or lack of it) with which their output is produced. Indeed, the fact that hospitals and nursing homes are typically reimbursed on a full-cost recovery basis (including full absorption of any overhead), is often suggested as the reason why that sector would probably remain a model of inefficiency even if the professionalization of hospital personnel no longer imposed institutional constraints on the effective use of health manpower.

By contrast, if health-care providers are paid for their services on a prepaid basis (systems E and F in Table 3), their net economic position varies inversely with their operating costs. Unless any deficit incurred in year t can be automatically loaded onto the prepayment premium for year $t + 1$—a situation that might obtain in a region whose entire health-care system is in the hands of a single prepaid health insurance plan—health-care providers operating under the prepayment mode have every incentive to minimize the cost of the services they render. This will be especially so if there is competition among providers so that the annual prepayment is not likely to contain a sizeable cushion for possible cost overruns. This inherent incentive to produce given services efficiently is one of the major benefits claimed for the prepayment mode.

The prepayment mode, however, is apt to force producers not only to produce given services efficiently, but also to search for the most efficient *mix* of services capable of maintaining the health status of patients at risk. It will be recalled that, in an earlier section, a clear distinction was drawn between efficiency in the production of *health services* and efficiency in the production of *health*. From that discussion it follows that it is not enough to confront health-care providers with incentives to produce specific services efficiently; that arrangement might still permit the delivery of an inefficient overall combination of health services. The great advantage inherent in prepayment plans such as Kaiser or HIP is that the financial incentives faced by these providers forces them to optimize over the entire range of medical services normally produced by a health-care system. The positive effect this optimization process appears to have on the mix of ambulatory and inpatient care has already been discussed. There is also some speculation that prepayment will, over the long run, lead to a desirable trade-off between preventive and therapeutic care. This rather powerful incentive toward global

efficiency in health maintenance is probably the single most important factor rendering prepaid comprehensive group practice so attractive to many health experts. The important point to keep in mind, however, is that it is the prepayment feature and not necessarily the group-practice setting that yields the desired efficiency gains in prepaid group practices.

Proposed Changes in the Financing of Health-Care Delivery

At the present time, virtually all of the financial arrangements shown in Table 3 exist side by side in the United States health-care sector. The great bulk of ambulatory care is being delivered under arrangement A; the bulk of inpatient care is delivered under arrangements B and D. Only a relatively small proportion of health services is now being rendered under arrangements E and F, i.e., prepaid comprehensive health plans.

Opinions concerning the most desirable financial arrangement for the American context differ sharply. It is probably fair to say that the medical profession as a whole continues to favor the fee-for-service mode, although there may not be a strong preference for either system A or B. Similarly, one gains the impression from the literature that the majority of health economists tend to favor a system under which consumers bear at least a part of the cost of the health services they consume, a notion that argues in favor of arrangements A, C, or E rather than B, D, or F (most economists would, however, argue also in favor of an upper limit of annual health-care costs borne directly by consumers).

A rather popular notion, held also by many knowledgeable students of health systems, is that the delivery of medical care had best not be polluted by the intrusion of financial incentives in the first place; that is, that both the producers and the consumers of medical care ought to make their decisions without reference to financial variables and purely on the basis of medical criteria. On this view, the economist's analogies between consumer goods in general and medical care are rejected as inappropriate and misleading; indeed, the very use of the terms "consumer" and "consumption" are held to be ludicrous. Similarly, it is held that physicians and other health-care providers are at their professional best when they can pursue their craft strictly as scientists and healers and not as classical entrepreneurs or modern business managers. The financial arrangement most compatible with this view would be the one under

which patients are fully covered by comprehensive health insurance, physicians work on a salaried basis, and other providers operate under, say, a budgetary system, preferably with somewhat flexible budget constraints. Such a system would be one version of arrangement D in Table 3. It has to be admitted that an arrangement of this sort has a certain intuitive appeal. Within the American context, however, the scheme strikes at least this author as rather utopian, resting as it does on somewhat utopian assumptions. In any case, it may be doubted that such a system will soon emerge on the North American continent.

The considerable advantage of the prepayment mode—systems E and F in Table 3—has already been indicated at some length. The long-established prepaid comprehensive group-practice plans (PGP) operating in various parts of the country are, of course, textbook models of such arrangements. Although there appears to have been early opposition to PGP's in this country, the familiar concept has recently been rediscovered, so to speak, and repackaged under the new label of Health Maintenance Organization (HMO). Under its new label, the concept is now being marketed with remarkable vigor by both the current Administration and the loyal opposition (the concept is a prominent feature of the national health insurance scheme proposed by Senator Edward Kennedy). The only new feature of the HMO concept is that, in the minds of at least some authors (e.g., Ellwood, 1971) the *prepayment* feature is no longer viewed as inseparable from the *group-practice* mode.

Careful thought has led some authors to wonder whether the HMO concepts is, in fact, likely to be the long-sought panacea for the shortcomings of the existing health-care system (see the rather critical review of the concept by Klarman, 1971). Unless the prepaid mode is made compulsory for all residents, individual HMO's may be plagued by adverse selection, with low-risk consumers opting for cheaper coverage under the traditional fee-for-service system. In addition, consumers must be given assurance that the HMO's natural incentive to minimize costs will not operate at the expense of quality. Finally, if the prepayment mode is coupled with a group-practice plan that operates its own inpatient facilities, the viability of the HMO is contingent on a rather substantial enrollment (between 20,000 and 50,000 at a minimum) which, in turn, may preclude the simultaneous existence of competing HMO's in

many regions. There may, therefore, arise a need to regulate HMO's as quasi-public utilities, a requirement that raises enormous problems in its own right. A more penetrating discussion of these and other economic problems associated with the HMO concept clearly goes beyond the limits of the present paper (for a fuller discussion of these issues, see Reinhardt, 1972b). Suffice it to say that the theoretically attractive notion of prepayment for medical service may not work in quite the expected manner in practice.

If one were forced to predict the probable evolution of the United States health system during the next decade or so, one would be safest to place one's bet on system B in Table 3, that is, on a system under which consumers receive medical services at zero or negligible marginal (out-of-pocket) costs, while the bulk of providers—and certainly medical practitioners—are reimbursed on a fee-for-service basis. It is perhaps no accident that the Canadian health system has already moved to that arrangement: within the North American context, that system simply has enormous political appeal. And even if in some intervening years the American population were to be covered by a national health insurance scheme with high deductibles and/or coinsurance rates, such a system, responding as it would to continued political pressure, would undoubtedly evolve in the direction of successively lower cost-sharing on the part of consumers. The utilization pattern and delivery system emerging under the Canadian health system may therefore be taken as a preview of probable future developments in the United States. Of all conceivable financial arrangements that may accompany the delivery of health care, it is the one system least likely to induce efficiency in the production of medical services and to encourage economy in their use.

Summary and Conclusions

Even a cursory reading of the literature on health care, both academic and popular, convinces one that there is widespread disenchantment with the nation's existing health-care system. It is argued that the present system fails to deliver what is needed, where it is needed, and at the right time. Furthermore, there is persuasive empirical evidence that such services as are made available to consumers are often produced inefficiently and are therefore unnecessarily costly. As a result, even those individuals who can afford the price

of normal health-maintenance services are constantly exposed to the risk of financial ruin through illness.

It would, of course, not be terribly difficult to protect all American residents from the financial risk of major illness, and such protection is, indeed, the main objective of the health-insurance proposals now before Congress. Many observers of the health-care systems believe, however, that any major national health insurance scheme would be doomed from the outset unless the health-care *delivery system* were first rationalized. The object of this rationalization would be, first, to lower the cost of medical care (or at least to reduce the upward pressure on medical-care prices) and, second, to enhance what is loosely referred to as the "overall quality of health care available to Americans."

Those whose task it is to fashion a more efficient provider structure out of the existing system can choose from a great variety of reform proposals. These proposals range all the way from marginal changes within the existing framework to a complete abandonment of the present, market-oriented structure in favor of a centrally directed system.

If one were forced to distill some common elements from the more ambitious reform proposals, one would have to cite at least the following three measures:

1. A consolidation of the existing agglomeration of small, more or less independent provider facilities into a coordinated network of larger units;
2. The substitution of relatively less expensive medical inputs (e.g., paramedical personnel) for more expensive ones (e.g., physicians) in the production of health care; and
3. A shift away from the traditional fee-for-service arrangement to a system of prepaid capitation fees for comprehensive health services.

The first and second of these measures are directed at the *technology* of health-care production. It is hoped that by moving in these directions, it will be possible to harness the economic benefits from as yet unexploited economies of scale and to enhance the productivity of the nation's most expensive and scarcest medical resource: the highly trained medical practitioner. The empirical evidence on

the presence of scale economies in health-care production is actually not very encouraging; one has the impression that proposals for a consolidation of provider facilities are argued as much on the basis of faith as on the basis of firm empirical support. The potential contribution of factor substitution, on the other hand, is far more obvious at this time. All available empirical evidence bearing on this issue suggests that this sort of substitution is probably the most effective way to arrest the steep secular increase in the cost of medical care. And it is a change that can be effected even within the existing framework of provider facilities.

The third measure listed above—i.e., the shift from fee-for-service payment to prepaid capitation—has to do with the financial incentives confronting both the consumers and the providers of health services. It is argued by the proponents of such a shift that prepayment for comprehensive health care will force providers of health services to seek the utmost efficiency in the production of these services, as well as to render consumers a mix of services that is capable of maintaining the consumers' health at minimum overall costs. At the same time, it is argued, prepayment for health services removes the financial disincentives normally faced by consumers each time a medical service is demanded. It is hoped that the removal of this disincentive will encourage consumers to avail themselves of medical attention prior to or in the early stages of a medical condition when treatment costs are relatively low. In the long run, the overall cost of health maintenance (at given levels of health status) should therefore be relatively lower.

Prepaid multi-specialty group practices, providing comprehensive care on their premises or in affiliated hospitals, are often said to incorporate all of these reforms simultaneously, and a system dominated by this form of provider facility is therefore widely viewed as the logical alternative to the existing American health system. This idea quite obviously lies also at the heart of the Administration's concept of a Health Maintenance Organization (HMO). To quote from the President's message to Congress on a "National Health Strategy" (1971:3016–3017):

> In recent years, a new method for delivering health services has achieved growing respect. *This new approach has two essential attributes.* It brings together a comprehensive range of medical services in a *single organization* so that a patient is assured of convenient access to all of them. And it provides needed services for a

> fixed contract fee which is *paid in advance* by all subscribers. . . . Some seven million Americans are now enrolled in HMO's—and the number is growing. Studies show that they are receiving high quality care at a significantly lower cost—as much as one-fourth to one-third lower than traditional care in some areas. [Italics added.]

That by "single organization" is meant a multi-specialty group practice may be inferred from the President's subsequent analogy between HMO's and the supermarket concept in retailing. It is also clear that the comparative cost studies referred to in the message consist in the main of studies in which the performance of the Kaiser Plan in California or the Health Insurance Plan (HIP) of Greater New York—both providing their ambulatory care on the premises of large-scale multi-specialty groups—was compared to health insurance plans—mainly Blue Cross-Blue Shield plans—relying on the more traditional fee-for-service providers (often solo practitioners). Such comparisons, however, must be viewed with some caution, for any of the following reasons:

1. It is not clear that one can legitimately project the experience of the particular populations covered by Kaiser or HIP onto the entire United States population. In particular, it should be noted that the population density in areas now covered by prepaid group plans tends to be much higher than that for the United States as a whole.

2. The often lower utilization rates reported by the prepaid group plans may reflect existing *supply constraints* more than deliberate choice. For example, the Kaiser plan provides both fewer physicians and fewer hospital beds per capita than are available to the United States population as a whole. The interesting question is whether these lower rates of factor inputs are the *result* of lower utilization rates, or whether they have been predetermined by administrative decision and *caused* the lower utilization rates. Since in the hospital sector, in particular, the supply of beds is often said to create its own demand for beds, it is quite conceivable that in an environment with less stringent supply constraints a Kaiser-type plan will generate much higher per-capita costs than are now observed for that organization.

3. It is possible that the consumers now enrolling in prepaid group plans and the physicians practicing there are somehow *atypical* of the United States population as a whole—one more reason why the cost studies referred to above may not be reliable.
4. It is never clear from these cost comparisons how much of the observed cost savings under the prepaid group-practice mode are due strictly to the prepayment feature and how much is due to the economies of scale attributed to the group-practice technology.

The latter of these points is of particular importance for, as was mentioned earlier in this essay, the proponents of the prepaid group-practice mode sometimes overlook the fact that the financial arrangements accompanying the delivery of health care can vary quite independently from the technology of health-care production. Solo medical practice, for example, can go hand in hand with prepaid capitation reimbursement, as is the case in Great Britain. Similarly, multi-specialty group practice can thrive under the fee-for-service system, as is predominantly the case where such practices exist in Canada or in the United States. A priori, there is therefore no compelling reason why a Health Maintenance Organization should not offer its subscribers all ambulatory care through a battery of solo practitioners, each of whom is paid in some manner by the HMO. The only crucial requirements are (1) that consumers themselves *prepay* for their care, and (2) that the governing board of the HMO is free from the control of physicians who deliver services on behalf of the HMO. These conditions would ensure, on the one hand, that the HMO is forced to have all necessary care produced as efficiently as possible and, on the other, that there would emerge a more competitive market for medical services, a market in which a well-informed HMO management—rather than numerous ill-informed individuals, often in ill health—would bargain with health-care producers. That degree of competition does, of course, not now exist in the health-care market.[8]

[8]As was mentioned earlier, if HMO's actually own the medical practices and hospital facilities they use, they may have to have such large enrollments as to preclude the existence of other provider facilities in a given region. In that case the HMO might well become a natural monopoly and eliminate even the very modest degree of competition now operating in the health-care market.

Perhaps because of its great intuitive appeal, the supposed advantage of group medical practice has long gone unchallenged among social scientists. The group-practice setting is assumed to generate significant economies of scale, to enhance the continuity of care, to economize on the patient's own time and, most important of all, to permit more effective peer supervision than is possible under a loosely connected network of small solo practices. These alleged benefits are clearly not trivial, and it is because of their presumed magnitude that the group-practice concept is often viewed as the *sine qua non* of the HMO.

Of late, however, at least some economists and sociologists have had second thoughts about the relative merits of the group-practice setting per se (Bailey, 1970a and 1970b; Freidson, 1970; Newhouse, 1973). First, it is pointed out, there is no conclusive empirical evidence that the production of ambulatory care is in fact characterized by economies of scale. Second, it is not clear to what extent a shift from solo to group practice would alter the behavior of medical practitioners. The proponents of group medical practice point to the exemplary behavior of physicians now working in Kaiser clinics or other prepaid groups, all the while overlooking the fact that these physicians are a highly select group that may not be, and most probably is not, representative of American physicians as a whole. The economic theory of groups in general suggests, for example, that since all benefits and costs in a group are shared equally by all of its members, each individual member tends to feel less compulsion to exercise economy than he would as an individual entrepreneur. The fact that Kaiser physicians do not appear to behave in this way surely does not invalidate the theory.

With respect to the benefits from peer supervision and its implication for standards of quality, there comes to mind Eliot Freidson's thought-provoking book on *Professional Dominance,* in which the author argues (1970:220–221):

> . . . when men work together in the same place, on the same terms, and with common work problems, they will develop a set of standards and procedures by which to judge and manage those problems, and they will discourage deviation from those standards. . . . [But merely] the fact of participating together in the same organized setting provides no assurance that the actual standards colleagues do agree upon and enforce will be either technically or socially adequate: the only assurance is that there will be standards of some

> kind and that they are likely to be narrower and better enforced than if the same men were scattered through a variety of settings . . . . [It] follows that considerable caution must be exercised in developing new forms of medical care organization. *The ideal picture of solo practice advanced by organized medicine should not be rejected out of hand merely because it does not reflect reality (or because it has been advanced by organized medicine): it represents a very flexible method of providing primary care that can, under some circumstances, be very satisfying to both physicians and patients.* [Italics added.]

Freidson's plea for flexibility serves as a fitting conclusion to this essay. If that conclusion appears somewhat agnostic, it has been correctly read. Surely one cannot come away from a survey of the literature on the health-care sector with the impression that the behavior of that sector is at all well understood at this time. And there is even more uncertainty concerning the ultimate, long-run effect any of the proposed changes in the existing system will have on the cost and quality of health care in this country (see also Reinhardt, 1972b).

The agnosticism evinced here is really also a plea for much more extensive empirical research than has hitherto been devoted to the organization of health-care delivery. Such research is methodologically difficult and, by virtue of its enormous data requirements, expensive. But in the long run it is likely to be cost-effective.

Uwe E. Reinhardt, PH.D.
Assistant Professor of Economics and Public Affairs
Woodrow Wilson School
Princeton University
Princeton, N.J. 08540

References

American Medical Association

1964 Report of the Commission on the Cost of Medical Care, Vol. 1. Chicago: American Medical Association.

1971 Reference Data on the Profile of Medical Practice. Chicago: American Medical Association.

Andersen, Ronald, Bjorn Smedby, and Odin W. Anderson

1970 Medical Care in Sweden and the United States: A Comparative Analysis of Systems and Behavior. Chicago: Center for Health Administration Studies.

Anderson, Odin W.

1964 "Research in hospital use and expenditures." Journal of Chronic Diseases 15 (September):727–733.

Badgley, Robin F., and Samuel Wolfe

1967 Doctors' Strike: Medical Care and Conflict in Saskatchewan. New York: Atherton Press.

Bailey, Richard M.

1968 "A comparison of interests in solo and fee-for-service group practice in the San Francisco Bay areas." Bulletin of the New York Academy of Medicine 44 (November):1293–1303.

1970a "Philosophy, faith, fact and fiction in the production of medical services." Inquiry 7 (March):37–53.

1970b "Economies of scale in medical practice." In Klarman, Herbert E. (ed.), with the assistance of Helen H. Jaszi, Empirical Studies in Health Economics. Proceedings of the Second Conference on the Economics of Health. Baltimore: The Johns Hopkins Press.

Balfe, Bruce E.

1969 "A survey of group practices in the United States, 1965." Reprinted from Public Health Reports 84 (July):597–604.

Bashshur, Rashid L., and C. A. Metzner

1967 "Patterns of social differentiation between Community Health Association and Blue Cross-Blue Shield." Inquiry 4 (June):23–44.

Bashshur, Rashid L., Charles A. Metzner, and Carla Worden

1967 "Consumer satisfaction with group practice, the Cha case." American Journal of Public Health 57 (November):1991–1999.

Berki, Sylvester E.

1972 Hospital Economics. Lexington, Massachusetts: D. C. Heath & Co.

Berry, Ralph

1970 Prepayment Plans (Processed).

Boan, John A.

1966 Group Practice. Ottawa: Queen's Printer.

Carr, W. John

1970 "Economic efficiency in the allocation of hospital resources central planning vs. evolutionary development." In Klarman, Herbert E. (ed.), with the assistance of Helen H. Jaszi, Empirical Studies in Health Economics. Proceedings of the Second Conference on the Economics of Health. Baltimore: The Johns Hopkins Press.

Cashin, James A.

1970 The Impact of the Advent of Medicare on Hospital Costs. Hempstead, New York: Hofstra University Yearbook of Business, Series 7, Vol. 2.

Clarkson, Kenneth W.

1971 Property Rights, Institutional Constraints and Individual Behavior: An Application to Short Term General Hospitals. Unpublished Ph.D. dissertation, University of California at Los Angeles.

Clute, K. F.

1963 The General Practitioner: A Study of Medical Education and Practice in Ontario and Nova Scotia. Toronto: University of Toronto Press.

Cohen, Harry A.

1967 "Variations in cost among hospitals of different sizes." Southern Economic Journal 33 (January):355–366.

Committee on Costs of Health Services

1970 Task Force Reports on the Cost of Health Services in Canada, Vol. 1. Ottawa: Queen's Printer.

Davis, Karen

1969 A Theory of Economic Behavior in Non-Profit, Private Hospitals. Unpublished Ph.D. dissertation, Rice University, Houston, Texas.

Densen, Paul M., Eve Balamuth, and Sam Shapiro

1958 Prepaid Medical Care and Hospital Utilization. Chicago: American Hospital Association.

Densen, Paul M., E. W. Jones, Eve Balamuth, and Sam Shapiro

1960 "Prepaid medical care and hospital utilization in a dual choice situation." American Journal of Public Health 50 (November):1710–1726.

Densen, Paul M., Sam Shapiro, E. W. Jones, and I. Baldinger

1962 "Prepaid medical care and hospital utilization: comparison of a group practice and a self-insurance situation." Hospitals 36 (November 16):62–68, 138.

Dominion Bureau of Statistics

1971 Canada Year Book 1970–71. Ottawa: Information Canada.

Donabedian, Avedis

1966 "Evaluating the quality of medical care." Milbank Memorial Fund Quarterly 44 (July, Part 2):166–206.

1969 "An evaluation of prepaid group practice." Inquiry 6 (September): 3–27.

Donabedian, A., S. J. Axelrod, and Judith Agard

1970 Medical Care Chart Book. Ann Arbor: The University of Michigan School of Public Health, Bureau of Public Health Economics.

Dunlop, John T.

1965 "The capacity of the United States to provide and finance expanding health service." New York Academy of Medicine Bulletin 14, (December):1326–1327.

Eiseman, B., and F. C. Spencer

1965 "The occasional open-heart surgeon." Circulation 31 (February): 161–162.

Ellwood, Paul M., Jr.

1971 "Restructuring the health delivery system—will the health maintenance strategy work?" In Health Maintenance Organizations: A Reconfiguration of the Health Services System. Proceedings of the Thirteenth Annual Symposium on Hospital Affairs, Center for Health Administration Studies, University of Chicago (May): 2–11.

Fein, Rashi

1967 The Doctor Shortage: An Economic Diagnosis. Washington, D.C.: The Brookings Institution.

Feldstein, Martin S.

1968 Economic Analysis for Health Service Efficiency. Chicago: Markham Publishing Company.

1971 The Rising Cost of Hospital Care. Washington, D.C.: Information Resources Press.

Foundation on Employee Health, Medical Care and Welfare

1962 Family Medical Care Under Three Types of Health Insurance. New York: The Foundation on Employee Health, Medical Care and Welfare.

Freidson, Eliot

1961 Patients' Views of Medical Practice—A Study of Subscribers to a Prepaid Medical Plan in the Bronx. New York: Russell Sage Foundation.

1970 Professional Dominance: The Social Structure of Medical Care. New York: Atherton Press.

Fry, John

1970 Medicine in Three Societies: A Comparison of Medical Care in the USSR, USA and UK. New York: American Elsevier Publishing Co.

Garbarino, Joseph W.

1960 Health Plans and Collective Bargaining. Berkeley: The University of California Press.

Ginzberg, Eli, with Miriam Ostow

1969 Men, Money, and Medicine. New York and London: Columbia University Press.

Glaser, William A.

1970 Paying the Doctor: Systems of Remuneration and Their Effects. Baltimore: The Johns Hopkins Press.

Goldman, Franz

1946 "Potentialities of group practice of medicine." The Connecticut State Medical Journal 10 (April):289–294.

Greenfield, Harry I., with the assistance of Carol A. Brown

1969 Allied Health Manpower: Trends and Prospects. New York: Columbia University Press.

Henderson, M. B.

1966 "Federal employees health benefits program, II: role of the group practice prepayment plans." American Journal of Public Health 56 (January):54–57.

Hopkins, Carl E. (ed.)

1969 Outcomes Conference I-II: Methodology of Identifying, Measuring and Evaluating Outcomes of Health Service Programs, Systems and Subsystems. Washington, D.C.: Health Services and Mental Health Administration.

Hutchinson, G. B.

1960 "Evaluation of preventive services." Journal of Chronic Diseases 11 (May):497–508.

Ingbar, Mary L., and Lester D. Taylor

1968 Hospital Costs in Massachusetts. Cambridge, Massachusetts: Harvard University Press.

Joseph, Hyman

1971 "Empirical research on the demand for health care." Inquiry 8 (March):61–71.

Kehrer, Barbara H.

1970 The Nursing Shortage and Public Policy: An Economic Analysis of the Demand for Hospital Nurses in Connecticut. Unpublished Ph.D. dissertation, Yale University.

Klarman, Herbert E.

1965 The Economics of Health. New York: Columbia University Press.

1968 "Effect of prepaid group practice on hospital use." Public Health Reports 78 (November):955–965.

1969 "Approaches to moderating the increases in medical care costs." Medical Care 7 (May, June):175–190.

1971 "Analysis of the HMO proposal—its assumptions, implications, and prospects." In Health Maintenance Organizations: A Reconfiguration of the Health Services System. Proceedings of the Thirteenth Annual Symposium on Hospital Affairs, Center for Health Administration Studies, University of Chicago (May):24–38.

Lando, Mordechai E.

1971 "A comparison of the military and civilian health systems." Inquiry 8 (June):56–61.

Lave, Judith R.

1966 "A review of methods used to study hospital costs." Inquiry 3 (May):57–81.

Lee, Maw Lin

1971 "A conspicuous production theory of hospital behavior." Southern Economic Journal 38 (July):48–58.

Lerner, Monroe

1961 Hospital Use by Diagnosis: A Comparison of Two Experiences. New York: Health Information Foundation, Research Series 19.

Long, M. F., and Paul J. Feldstein

1967 "Economics of hospital systems: peak loads and regional coordination." American Economic Review 57 (May): 119–129.

MacColl, William A.

1966 Group Practices and Prepayment of Medical Care. Washington, D.C.: Public Affairs Press.

Makover, Henry B.

1951 "The quality of medical care." American Journal of Public Health 41 (July):824–832.

Mann, Judith K., and Donald E. Yett

1968 "The analysis of hospital costs: a review article." Journal of Business 41 (April):191–202.

McCaffree, Kenneth M., and Harold F. Newman

1968 "Prepayment of drug costs under a group practice prepayment plan." American Journal of Public Health 58 (July): 1212–1218.

McElrath, D. C.

1961 "Perspective and participation of physicians in prepaid group practice." American Sociological Review 26 (August):596–607.

McNamara, Mary E., and Clifford Todd

1970 "A survey of group practice in the United States, 1969." Public Health Reports 60 (July):1303–1313.

McNerney, Walter J., and study staff

1962 Hospital and Medical Economics: A Study of Population, Services, Costs, Methods of Payment, and Controls, Vols. 1 and 2. Chicago: Hospital Research and Educational Trust.

Monsma, George N., Jr.

1970 "Marginal revenue and the demand for physicians' services." In Klarman, Herbert E. (ed.), with the assistance of Helen H. Jaszi, Empirical Studies in Health Economics. Proceedings of the Second Conference on the Economics of Health. Baltimore: The Johns Hopkins Press.

Morehead, Mildred A.

1967 "The medical audit as an operational tool." American Journal of Public Health 57 (September):1643–1656.

Myers, Robert J.

1970 "A government insider denounces medical fee controls." Reprint of a speech before Oklahoma State Medical Association. Medical Economics 47 (July 20):218–233.

Newhouse, Joseph P.

1973 "The economics of group practice." Journal of Human Resources 8 (Winter):37–56.

Nixon, Richard

1971 President's Message on Health and Hospitalization. Congressional Record, 92nd Congress, First Session, 117, Part 3: 3015–3021. Washington, D.C.: U.S. Government Printing Office.

Pauly, Mark V.

1970 "Efficiency, incentives and reimbursement for health care." Inquiry 7 (March):114–131.

Peterson, Osler L.

1963 "Medical care: its social and organizational aspects." The New England Journal of Medicine 269, Number 23 (December):1238–1245.

Pondy, L. R.

1971 Physician assistants' productivity: an interim report. Durham, North Carolina: Duke University School of Public Administration (January).

President's Commission on Heart Disease, Cancer and Stroke

1965 Report to the President: A National Program to Conquer Heart Disease, Cancer and Stroke, Vol. 2. Washington, D.C.: U.S. Government Printing Office.

Reder, Melvin W.

1969 "Some problems in the measurement of productivity in the medical care industry." In Fuchs, V. R. (ed.), Production and Productivity in the Service Industries. National Bureau of Economic Research, Studies in Income and Wealth, Vol. 34. New York: Columbia University Press.

Reinhardt, Uwe E.

1970 An Economic Analysis of Physicians' Practices. Unpublished Ph.D. dissertation, Yale University.

1972a "A production function for physicians' services." Review of Economics and Statistics 54 (February):55–66.

1972b On the Economic Characteristics of the Health Maintenance Organization. (Processed)

Ro, Kong Kyun

1968 "Determinants of hospital costs." Yale Economic Essays 18 (Fall): 185–257.

Rutstein, David D.

1966 "At the turn of the next century." In Knowles, John H. (ed.), Hospitals, Doctors, and the Public Interest. Cambridge, Massachusetts: Harvard University Press.

Saskatchewan Medical Care Insurance Commission

1969 Annual Report. Regina, Saskatchewan: Saskatchewan Medical Care Insurance Commission.

Schwartz, William B.

1970 "Medicine and the computer: the promise and problems of change." New England Journal of Medicine 283 (December 3): 1257–1264.

1971 Hearings before the Subcommittee on Health of the Committee on Labor and Public Welfare, U.S. Senate, 92nd Congress, First Session, on Examination of the Health Care Crisis in America, Part 3: 442–463. Washington, D.C.: U.S. Government Printing Office.

Shapiro, Sam

1967 "End result measurements of quality of medical care." Milbank Memorial Fund Quarterly 45 (April):7–30.

Shapiro, Sam, Louis Weiner, and Paul M. Densen

1958 "Comparison of prematurity and perinatal mortality in a general population and in the population of a prepaid group practice, medical care plan." American Journal of Public Health 48 (February):170–187.

Smith, Kenneth R., Marianne Miller, and Frederick L. Golladay

1972 "An analysis of the optimal use of inputs in the production of medical services." Journal of Human Resources 7 (Spring):208–225.

Somers, Anne R.

1971 Health Care in Transition: Directions for the Future. Chicago: Hospital Research and Educational Trust.

U.S. Bureau of the Census

1971 Statistical Abstract of the United States, 1971. Washington, D.C.: U.S. Government Printing Office.

—Department of Health, Education, and Welfare

1967 Medical Care Prices. Report to the President. Washington, D.C.: U.S. Government Printing Office.

—National Advisory Commission on Health Manpower

1967 Report of the National Advisory Commission on Health Manpower. 2 Vols. Washington, D.C.: U.S. Government Printing Office.

Weinerman, E. Richard

1964 "Patients' perceptions of group medical care: a review and analysis of studies on choice and utilization of prepaid group practice plans." American Journal of Public Health 54 (June):880–889.

Weiss, Jeffrey H.

1966 The Changing Job Structure of Health Manpower. Unpublished Ph.D. dissertation, Harvard University.

Wolfman, Burton

1961 "Medical expenses and choice of plans: a case study." Monthly Labor Review 84 (November):1186–1190.

Yankauer, Alfred, John P. Connelly, and Jacob J. Feldman

1970 "Physician productivity in the delivery of ambulatory care: some findings from a survey of pediatricians." Medical Care 8 (January-February):35–46.