This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Production and Productivity in the Service Industries

Volume Author/Editor: Victor R. Fuchs, ed.

Volume Publisher: NBER

Volume ISBN: 0-870-14489-8

Volume URL: http://www.nber.org/books/fuch69-1

Publication Date: 1969

Chapter Title: Some Problems in the Measurement of Productivity in the Medical Care Industry

Chapter Author: M. W. Reder

Chapter URL: http://www.nber.org/chapters/c1201

Chapter pages in book: (p. 95 - 153)

Some Problems in the Measurement of Productivity in the Medical Care Industry

M. W. REDER

STANFORD UNIVERSITY

GIVEN the operational definitions of inputs and outputs, the conceptual problems of measuring productivity change are much the same in one industry as in another. The present paper is not addressed to these generic problems, except incidentally, but is focused on the problems of conceptual specification that arise in studying this particular industry.

Medical care involves more than the cure of ailments; indeed, considering the importance of psychosomatic factors in the genesis of disease and of socio-economic factors in the genesis of psychological disturbance, most household consumption could be treated as preventive medicine. But however defensible such a classification might be, it is not consonant with either individual or social resource allocation procedures. Despite mixed motives, confusion of purpose, and a high propensity to convert erstwhile pleasures into current medical necessities, we still have fairly definite standards as to what we will treat as medical care expenditure. These standards are manifested in what is covered by publicly sponsored medical insurance programs or privately financed medical insurance plans; by what the Bureau of Internal Revenue will allow as a medical expense deduction for income tax purposes, and so forth.

In essence, a service will usually be considered as medical care if

NOTE: This paper was written as part of a project on the Supply and Demand of Medical Personnel, sponsored by the Ford Foundation.

it is provided by or under the direction of a licensed physician or if its provision by anyone else would be cause for prosecution for practicing medicine without a license.[1] Services recommended as beneficial by a physician may or may not be considered as medical care; e.g., ocean voyages, frequent swimming in warm water, and so forth, may be recommended by a physician, but are not always considered as medical care by the Internal Revenue Service and hardly ever by the keepers of the National Accounts.

For simplicity, let us make the boundaries of the medical care industry coterminous with the coverage of medical care insurance or publicly provided comprehensive medical care service—whichever is wider. Clearly, in most advanced countries there is a consensus that what is considered medical care is to be favored (subsidized) relative to most other objects of consumer expenditure. It is necessary, therefore, that there be some fairly sharp distinction between what is to qualify for subsidy and what is not. Similarly, purveyors of medical care insurance must sharply distinguish between those items of expenditure for which they will pay and those for which they will not. Normally, to qualify as medical care, an object of expenditure must have either a significant element of immediate physical unpleasantness about it or involve the requisition by a (presumably) disinterested third party, usually a physician. Failing third party control, there is serious risk that claims for compensation would become so frequent as to drive insurance premia to unattractively high levels in order to cover costs.[2]

An obvious locus of control over utilization of medical services is the physician. Indeed, the physician's claim to complete control over patient treatment makes it almost impossible to locate the control elsewhere. On this basis, one could define medical care expenditure

[1] The aftermath of an illness may involve a prolonged period of care (e.g., in a nursing home), which may be physician-prescribed but not physician-controlled. During such periods, distinguishing between ordinary consumption and medical care involves many arbitrary decisions.

[2] Of course, insurance can be made "actuarially fair" at any level of claims, if the level of premia is appropriately adjusted. However, insurance is usually attractive (to those insured) only when infrequent claims permit low premia relative to the average size of claim. "Overuse," therefore, threatens the viability of insurance systems leading to tightened controls as well as high premia. Similarly, overuse of publicly provided or subsidized services leads to their curtailment, adulteration, or both.

as that which is ordered by a physician for the benefit of a patient. This view of medical care implies a very sharp distinction between the well individual and the sick (ideally, hospitalized) one; none of the former's expenditure is controlled by the physician (and therefore none goes for medical care) while all of the current consumption of the hospitalized individual is medical care. Obviously much thinking about the economics of medical care, past and present, reflects this dichotomy.

But an increasingly important body of medical thought recognizes the impact of general consumption patterns (e.g., housing) upon health. Many physicians, consciously or otherwise, have begun to adjust treatment patterns to facilitate substitution of costless (to the patient) health-enhancing inputs for ordinary consumption. For example, physicians will sometimes hospitalize patients primarily because the home environment is deleterious (e.g., the patient will walk too many stairs; be exposed to an emotionally disturbing situation or to a bad diet) rather than because he "needs" hospital services. The extent to which this is done appears to vary with the availability of hospital beds,[3] which means, among other things, that increases in hospital productivity leading to a shorter hospital stay for a given medical condition may be obscured by an offsetting increase in the consumption of the hotel services rendered by hospitals. That is, increases in productivity in hospitals may go unrecorded because of an expansion of the range of services recorded as medical care. As will be seen, this possibility creates measurement difficulties in several contexts.

Nevertheless, considerations of public policy require that we maintain the distinction between medical care and other consumer expenditures;[4] and the policy considerations that impel maintenance of this distinction, suggest where it be drawn. Medical care expendi-

[3] There is a substantial body of evidence supporting this statement. Numerous citations have been made in a previous paper, "Some Problems in the Economics of Hospital Care," *American Economic Review,* May 1965, pp. 472–480, footnotes 5–15. M. S. Feldstein, *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service,* Amsterdam, 1967, pp. 193–222 and "An Aggregative Model of the Health Care Sector" (mimeographed), Oxford University, 1967, presents evidence, for both the U.K. and the U.S., that utilization of hospital beds is increased by an increase in supply.

[4] Because it is public policy to subsidize the former, but not the latter.

ture is that for which medical care insurance will pay (or at least make a contribution). Obviously, medical care insurance policies vary in coverage; for the purpose of this definition, consider the broadest type of coverage that is generally available.[5]

This definition of medical care expenditure suggests correlative measures of price change and productivity. If medical care is that which can be purchased by means of medical care insurance, then its "price" varies proportionately with the price of such insurance.[6] That is, for a person of given age and health, the mathematical expectation of expenditure on medical care will vary proportionately with the premium of a comprehensive medical care insurance policy.[7] Changes in resource productivity of medical care may then be measured as the reciprocal of the change in the ratio of an index of medical care insurance premia to an index of prices of inputs used in producing medical care.[8]

[5] For the purpose of this definition, the fraction of the cost borne by the insurer is irrelevant.

[6] Harold Watts has asked "if medical care is that which can be purchased by means of medical care insurance, isn't life that which can be purchased with life insurance?" Humor aside, the question raises an important issue: is insurance offered in real or in money (nominal) terms? Life insurance, like most other kinds, is really a promise to pay money as compensation for the occurrence of an undesired event. Medical care insurance may take the form of cash payments that serve as an offset to payments made for medical care or it may involve direct provision of care (as in a prepayment plan).

Direct provision of care under complete prepayment is real (medical care) insurance; payment of the insured's medical bills where the amount of insurance benefit is an increasing function of the size of the bill is also real insurance, with coinsurance aspects. Where the amount of the insurance benefit is fixed in money terms (independent of the dollar magnitude of the medical bills), insurance is nominal. Obviously, a given policy may be a combination of real and nominal (money) insurance. Insurance premia are assumed to be paid in money terms, though real premia are imaginable.

Where medical care insurance is real, the insurer must be concerned with prices of the goods and services he must either provide or pay for. As the quality or the amenities of care are likely to vary with its price, the insurer is also involved in matters regarding the kind of care to be provided. This brings the third party (the insurer) into the market for medical care as a party interested in minimization of his own liability. This possibility obviously cannot arise in the case of nominal insurance such as is afforded by life insurance.

[7] This statement assumes that the insuring industry is in long-run competitive equilibrium at all relevant moments and that the insurance covers all expenditure for medical care (i.e., the insurance is comprehensive and has no coinsurance features).

[8] Productivity change is measured here as the ratio of the relative change in the price of a unit of output to the relative change in the (weighted) average price

To define a price index of medical care in terms of the cost of medical care insurance differs from the procedure used in constructing the medical care component (M) of the Consumer Price Index; this component is essentially an index of input prices.[9] The medical care component can take no account of the effect of changed efficiency on the cost of medical care nor can it take account of changes in the "quality" of care. The "cost of insurance" index does reflect changes in the efficiency of producing medical care of given quality. Its major shortcoming, shared with M or any conceivable index, is that of distinguishing between price changes due to changing costs of producing an output unit of given quality, and those attributable to changes in the quality of the unit.

In order to have a meaningful measure of productivity change, it is essential to define a unit of output so that it can be measured independently of changes in the input units. To measure medical care in hours of physician time, hospital bed-days used, etc., or some composite thereof, precludes the possibility of detecting changes in the efficiency (greater or less output per unit of input) with which the care is produced. One way to circumvent this difficulty is to define medical care ostensively; i.e., medical care is that which is provided for an individual who has membership in (say) a particular comprehensive prepaid medical care plan.[10] In a comprehensive care plan, quantity of care is all that is "medically" needed as determined by the attending physician. (The complicated physician-patient interaction that actually determines quantity of care under any administrative system may be ignored for the moment; it is discussed in the next section.) Quality of care is designated as that which is delivered under

of the inputs used in its production. This implies neglect of changes in the ratio of average cost to price in the insuring industry.

[9] M is a weighted average of prices of services of doctors, nurses, drugs, appliances, hospital room rates, and other goods and services (inputs) used in producing medical care. The construction of the medical care component of the CPI is described in detail by E. A. Langford, "Medical Care in the Consumer Price Index, 1936–56," *Monthly Labor Review,* September 1957, pp. 1053–1058; and E. D. Hoover, "The CPI and the Problems of Quality Change," *Monthly Labor Review,* November 1961, pp. 1175–1185.

[10] I am here defining output of medical care as that *available* to a subscriber of a medical care plan, not what is used. This slurs over the important problem of the unpaid inputs in medical care, such as the time of the patient and his family, which will be discussed later. For the moment assume an individual's use of medical care is determined by his physician.

the plan, and changes in quality are measured by various performance measures as discussed below.

Changes in the quantity of care supplied are measured as follows: for each member of a given "treatment-need" class, defined by age, sex, and other relevant variables, the average inputs used in a base period are computed. Each person in the plan is weighted by the value of "inputs used per individual" in his treatment-need class, and the weighted sum computed.[11] The weighted number of individuals enrolled in the plan is the measure of output at specified quality, and changes in deflated value of input per unit of output (so defined) measure variations in productivity.

The critical assumption in this hypothetical procedure is that quality of care is unchanged. How can quality of care be measured? One method is as follows: it will be generally agreed that among the objectives of medical care are (1) lower age-sex specific mortality rates; (2) lower rates of undetected illness; (3) lower rates of improper treatment of specific ailments. It is, at least in principle, possible to compare different groups of subscribers (standardized for age and sex) with different medical care plans—or the same plan in different time periods—in respect of each of these three characteristics, and to rank the quality of health care received in terms of the results.

There are serious, but not insuperable difficulties to carrying out such a program in practice. Comparisons of mortality rates present no special problem. Comparison of different rates of undiagnosed ailments requires "medical audits" of sample groups of patients by a team of "experts"; such audits have been made, but notoriously there is resistance to such procedures.[12] Deciding upon the appropriateness

[11] The appropriate weighting or standardization techniques are well-known, and present no greater difficulties here than in most other uses.

[12] Among such "audits" that have been made are O. L. Peterson, *et al.*, "An Analytical Study of North Carolina General Practice, 1953–54," *Journal of Medical Education,* December 1956, Part 2; K. F. Clute, *The General Practitioner,* Toronto, 1963; and C. C. Jungfer and J. M. Last, "Clinical Performance in Australian General Practice," *Medical Care,* April–June 1964, pp. 71–83. Critical analyses of hospital admissions, dismissals, and treatment may be found in "Research in Hospital Use: Progress and Problems," Conference Report, U.S. Department of Health, Education and Welfare, Public Health Service (*Public Health Service Publication No. 930-E-1*), Washington, D.C., 1962. An excellent review of the literature on the evaluation of quality of medical care is Avedis Donabedian, "Evaluating the Quality of Medical Care," *Milbank Memorial Fund Quarterly,* July 1966, pp. 166–206; this source contains an extensive bibliography.

of courses of treatment requires examination of patient records and (on a sample basis) systematic scrutiny of treatment of individual patients from initial complaint (or detection of illness on a checkup) until completion. Here again, some pioneering work has been done,[13] but medical auditing is in its infancy.

Other important aspects of quality of health care can be readily monitored; patient satisfaction would surely be one. I claim no more than relevance for the aspects mentioned. It is probable that many of the interesting aspects will reflect a certain penumbra of professional disagreement as to which therapeutic techniques are superior. In the absence of strong consensus as to quality ranking of treatment methods or diagnostic procedures, there is little point to comparing the quality of treatment afforded different groups of persons. However, on a large number of important matters there is strong agreement as to what is an acceptable procedure and what is not.[14]

For purposes of productivity measurement, a dichotomy between acceptable and unacceptable procedures may be all that is needed. What we are trying to do is to compare ratios of outputs to inputs for different groups of individuals, and to take account of possible variations in quality that may be associated with differences in these ratios. The more sensitive the quality scales, the better; but for the economist the sensitivity of these scales is a datum. In the absence of unambiguous evidence of countervailing quality change, it is generally agreed that changes in methods of production that increase productivity are good—and vice-versa. But variations in quality of output, when they can be detected, serve as notice that increases in measured productivity may not have resulted from improved productive technique alone, and that a problem of relative valuation of quality change vis-a-vis pecuniary cost has arisen.[15]

To describe changes in productivity that accompany changes in quality of medical care I would recommend, as a first step, that the

[13] For example, Clute, *op. cit.*, Chapters 2, 16, and 17.

[14] For an illustration of this point see Peterson *et al.*, *op. cit.*, pp. 18–48, and Clute, *op. cit.*, Chapters 16 and 17.

[15] In an interesting paper, "Productivity and the Price of Medical Services," soon to be published in the *Journal of Political Economy*, Dr. Yoram Barzel attempts to measure changes in the cost of medical care by changes in premia for health insurance. His paper is in much the same spirit as the first section of this one, though the details are quite different.

productivity index and each dimension of quality of care be presented as distinct elements of a medical care vector. Consider the following hypothetical example: suppose that between years 0 and 1 productivity increased by 10 per cent; life expectancy at age 1 increased by six months; the number of cases (per 1,000 persons standardized) of undetected illness of specified kinds decreased by 3 per cent, but the percentage of improper treatments increased from 7 to 9 per cent of all cases. To reduce this vector to a scalar requires the postulation of a set of marginal rates of substitution (trade-offs) between the various elements; e.g., at a given rate of (1) annual change of productivity in medical care; (2) percentage of undetected illness; (3) percentage of improper treatments; (4) per capita income, and so forth, an *additional* 1 per cent increase in productivity per annum must be judged to be as desirable as, say, an *additional* ¼ per cent annual decline in undetected illness per 1,000 persons, or an *additional* ⅓ per cent annual decline in improper treatments.

Establishing these trade-offs is essential to any useful measurement of productivity in medical care. This is because it is so easy to increase (measured) productivity by adulterating product or (see below) appropriating unmeasured inputs.[16] And it is not impossible for individual investigators to establish some such set of trade-offs. But the process will involve difficult judgments of value (e.g., how much money it is worth to increase life expectancy by six months). Price data may be relevant to these judgments, but which prices should be used and how is likely to be a matter of substantial disagreement. Under the circumstances, measurers of productivity change may do well to present all the elements of the aforementioned vector and let each student combine them as he wishes.

The above procedure is oriented toward measuring the output of medical care used by members of comprehensive prepaid plans. But

[16] This possibility is not peculiar to the medical care industry, but arises wherever quality change is feasible. For an indication of the importance of "quality" change in automobile manufacturing, see F. K. Fisher, Z. Griliches, and C. Kaysen, "The Costs of Automobile Model Changes Since 1949," *Journal of Political Economy*, October 1962, pp. 433–451. J. L. Nicholson, "The Measurement of Quality Changes," *Economic Journal*, September 1967, pp. 512–530, offers a broad discussion of this subject and his footnotes contain numerous references to the relevant literature.

only a small part of the American population now is, or in the near future will be, enrolled in such plans. How then are we to measure output quantity for persons buying some or all of their medical care on a fee-for-service basis? The best, though very imperfect, answer that occurs to me is as follows: make elaborate periodic surveys of physician use by individuals (such as a national health survey) and treat base-year expenditures as the measure of base-year output. Follow up patient interviews with physician interviews to obtain reason or diagnosis for each contact. Typical cost of each specific illness can be computed as A. A. Scitovsky does (see below). Then relative output level in a given year may be measured as the relative number of illnesses treated, each illness weighted by an appropriate composite of its base- and given-year costs. Productivity change can be measured, as before, by dividing the index of relative outputs by an index of relative inputs. The input index can be computed from the data on the costs of specific illnesses. Quality changes can be measured by sample checkups on the physical condition of persons studied in the surveys of physician use.

At this point, economists may well protest at getting involved in the complicated problems of measuring quality of medical care and evaluating the social importance of its various aspects. Why not treat health care like other commodities and evaluate quality by price? If a "unit" of one type of care sells for twice as much as another at the same time and place, its production involves the use of twice the value of resources; [17] new types of care can be valued relative to old types via chain indexes as well (or badly) as new varieties of other products, and so forth.

I would not deny the possibility of treating medical care as "just another commodity" for purposes of measuring price and productivity changes. However, there are two peculiarities of the medical care product that are likely to make such treatment less than normally satisfactory. One is the prevalence of price discrimination that obscures the meaning both of "market price" and output unit.[18] A more

[17] The implicit assumption of competitive equilibrium in this statement may be challenged.

[18] Price discrimination in medical care is not as simple a matter as it is often thought to be. Granted that physicians may charge wealthy patients higher fees, it is not so obvious that they do not also give them more time, appointments on

fundamental difficulty is that for many purposes of social policy what is meant by quality of medical care is a technical medical judgment of efficacy in promoting "health," and not an inference from observed consumption patterns.

A useful though more limited measure of productivity variation may be inferred from Mrs. Scitovsky's study of the cost of specific illnesses.[19] By taking a number of specifically diagnosed illnesses, and following the recorded methods of treatment, she has been able to measure the specific inputs used in treating each illness at any given time, and to measure changes in inputs occurring over time.[20] Where effectiveness of treatment can be treated as unchanged, one may measure changes in productivity as the reciprocal of the (deflated) value of inputs per case. However, in some illnesses, technological progress has the effect of keeping people alive longer, but under continuing treatment. In other illnesses, new techniques promote more frequent recoveries, but also use more resources per case. In both of these types of illness, measurement of productivity requires allowance for quality change.

If we confine ourselves to cases where quality of treatment can be regarded as substantially unchanged over time (for the last decade, normal pregnancies and tonsillectomies might be examples), we can get an estimate of productivity change in medical care for quality-constant output. Such an estimate would be highly useful in determining whether, say, the organizational structure of the industry was making it less technologically progressive than others and, if so, which inputs were responsible.[21] One could also estimate the effect (on pro-

shorter notice, and other benefits. Also, physicians tend to give more time to "interesting patients" than to others without charging higher fees. The consequence of this is to make record keeping and data collection more than ordinarily difficult.

[19] A. A. Scitovsky, "Changes in the Costs of Treatment of Selected Illnesses, 1951–65," *American Economic Review,* December 1967, pp. 1182–1195.

[20] Actually what Mrs. Scitovsky has done is to measure changes in the money cost of treating specific illnesses. However, her procedure is such as to permit measurement of the inputs used at different times (at least to the extent of constructing an index), so that her data could be readily used for measurements of productivity.

[21] It is hazardous to infer a measure of quality-constant productivity change for all illnesses from its measure in those illnesses where change in quality of treatment has been negligible. Any such inference must either assume that increases in productive efficiency are the same in treatment of illnesses where quality of care is improving as in other illnesses, or make allowance for the differences.

ductivity) of changes in the mix of illnesses treated. Such measures could help answer many of the questions to which analysis of productivity change normally is applied. However, they would not enable us to measure those productivity changes that result in better health (i.e., more efficacious treatment) but not in greater economy of resources used per episode of illness.

Yet another approach to measuring output and productivity change is that of studying working days lost.[22] Clearly an important part of the economic loss from illness is the foregone output of disabled workers. A reduction in hours lost on account of illness (per labor force member, standardized for age and sex) indicates an improved quality of medical care. Evaluating the saved hours at their presumptive market value provides a quantitative measure of the quality improvement over the relevant period of time. However, this measure also has its shortcomings: (1) it is necessary to assume that the incentives and attitudes governing the allocation of time between work, leisure, and illness remain unchanged, or to allow for the effect of their variation; (2) it is also necessary to allow for the possible effect of earlier return to work (after an illness) on productive efficiency and (3) the measure takes no account of increased consumer satisfactions from better quality care; i.e., the value of keeping more people alive—or healthier—than before, though not at work.

Finally, let me note that there has been a considerable number of studies using a "straightforward" measurement of productivity change, with output defined as (deflated) value of sales, and little or no attempt to correct for quality variation.[23] These and other studies con-

[22] This has been done for U.S. Navy personnel in an interesting unpublished dissertation by M. E. Horton, at the University of Washington, 1966. B. A. Weisbrod, *Economics of Public Health*, Philadelphia, 1961, contains a serious effort to measure the pecuniary losses due to major diseases. Calculations of the type performed by Weisbrod can play an important role in measuring output of the medical care sector; i.e., reduced pecuniary losses from disease is one indicator of improved quality of medical care. H. E. Klarman, *The Economics of Health Care*, New York, 1965, p. 164, has suggested that the value of health as a consumer good may be inferred from the expenditure on nondisabling diseases of those not in the labor force.

[23] Two good recent discussions of productivity changes in medical care are Klarman, *ibid.*, pp. 149–157; and Rashi Fein, *The Doctor Shortage*, Washington, D.C., 1967, Chapter IV. Fein makes numerous references to more detailed studies, in particular to an unpublished dissertation by Jeffery Weiss, "The Changing Job

tain many points that merit discussion; some of them will be discussed below, but many others will not. The reason for bypassing them is to concentrate upon what is, at least to me, a more fundamental question: the relation of "industrial organization" to manpower productivity in medical care.

In this first section, I have attempted to make some suggestions for procedure in the measurement of productivity change in medical care. In so doing, I have assumed that quality of medical care could be defined in technical terms so that it would be meaningful for a public health expert to say, for example, that the care provided subscribers to the Health Insurance Plan of Greater New York was as good as that provided the average New Yorker under fee for service. This assumption is pragmatically useful for many purposes of public policy. But there is another aspect of medical care, the psychiatric-priestly (Rasputin?) aspect, that is not likely to be reflected in epidemiological data, no matter how good they are. This aspect has important effects on the economics of medical care and on productivity changes in particular; hence it is considered extensively in the remainder of the paper.

METHODS OF PAYMENT, INCENTIVES AND THE QUALITY-QUANTITY OF MEDICAL CARE

My colleague, Professor K. J. Arrow, has contended that medical care has many aspects that differentiate it from other commodities, and that these significantly affect the type of organizational framework most likely to approach Pareto optimality.[24] No doubt this is so, but to stress similarities or differences among commodities is mainly a matter of expository convenience. For the moment, it is the similarities between medical care and certain other commodities which are most important; later some differences will be discussed.

Structure of Health Manpower," Harvard University, 1966. M. S. Feldstein, *op. cit.*, presents an excellent cross-sectional analysis of cost and productivity differentials in British hospitals; and Alex Maurizi, "The Economics of the Dental Profession," unpublished doctoral dissertation, Stanford University, 1967, has made an important first step in explaining productivity differentials among dentists.

[24] K. J. Arrow, "Uncertainty and the Economics of Medical Care," *American Economic Review*, December 1963, pp. 941–973.

Methods of Payment

In principle all services rendered by humans (as distinguished from animals or inanimate capital) may be purchased under one of two types of contract: an employment contract or a sales contract.[25] Under an employment contract, the purchaser hires the services of an individual and assumes the responsibility for directing him so as to obtain the desired results.[26] Under a sales contract, the purchaser hires the supplier and allows the latter to do the job in the manner he considers appropriate.

For services whose satisfactory performance requires technical knowledge, it is impossible for the relatively ignorant buyer to supervise the seller. Such services include not only medicine, but also law, house construction, automobile repair and many others; sometimes even economic consulting. Clearly, in any of these cases a sales contract is in order. But what method of payment should be used? The supplier's incentive to perform will differ with the method of payment.

Consider the case of house construction: if the supplier (contractor) is paid a fixed fee, his incentive is to skimp on services rendered. If he is paid on a per hour basis, he has an incentive to reduce "effort" per hour. If he is paid strictly on the basis of "output," quality must be controlled lest he attempt to increase earnings by substituting quantity for quality. Any combination of the above involves weakening the incentive to err (from the buyer's viewpoint) in any one direction, but induces "errors" in some combination of all other directions.

Assume the person for whom the house is being built (the buyer) is attempting to maximize his expected utility and that, other things being equal, this depends upon the quality of the house, the date at which it is completed and the amount it costs him. Suppose the buyer is extremely averse to any reduction in the stock of his liquid asset holdings below some minimum level, and that as a result he seeks a contractor who agrees to construct the house (in accordance with specified plans) for a fixed price. This would seem to imply that the

[25] H. A. Simon, "A Formal Theory of the Employment Relationship," *Econometrica*, July 1951, pp. 293–305.

[26] It is assumed, implicitly, that the wage rate is high enough to make the employee anxious to avoid being discharged and therefore willing to accept direction.

risks of fluctuating input prices, random fluctuations in factor productivity, factor availabilities, etc., are borne by the contractor.

However, the contractor is also a utility maximizer; his maximization is subject to the constraints that the house be of minimum quality and be completed not later than some date. These constraints reflect, in uncertain proportions, the contractor's concern for his reputation, his fear of law suits or possible loss of license, and his ethics. Also assume, realistically, that both improving the quality of the house and advancing its date of completion have a positive and decreasing marginal utility to the contractor, but also have a positive and increasing marginal cost. Under these circumstances, an unanticipated rise in factor prices will lead to some decline in house-quality, and vice-versa.

Thus the risk of fluctuations in factor prices will be shared between the buyer and the contractor. How they are shared will depend crucially upon the elasticity of the contractor's marginal utility with respect to superior quality and prompter completion; the greater this elasticity, the smaller the loss to the buyer (in quality and speed of completion) from a given shift in the marginal cost of improving either one.[27] However, whatever change occurs in the final product, it will not—because of the fixed fee provision of the contract—affect the amount of money paid. Therefore, given that the contractor has some freedom to respond to unanticipated cost changes, he will do so in the quality or completion-date directions.

What is crucial in this argument is the implicit assumption that the contract does not completely determine what it is that the buyer is to obtain. In principle we might assume that it is possible to draw up building plans in sufficient detail and to specify a completion date, so that the contractor has no freedom of maneuver in respect of what he delivers.[28] But then what happens when in some respect he violates such a contract as, in an uncertain universe, he always will? Literal satisfaction of such a contract, including completion date, cannot be

[27] It also depends upon the buyer's bargaining skill; i.e., once a fixed fee contract is signed, important elements of a bilateral monopoly situation (between buyer and contractor) arise.

[28] Whether it is linguistically possible to specify quality completely is a deep question we may profitably avoid. It is sufficient to assume such specification too costly for frequent use.

required. However, forfeits and penalties may be assessed; the defaulting contractor may be compelled to relinquish his life, fortune, reputation or parts thereof. But in order to induce him voluntarily to agree to such rigid contract terms, the price offered for fulfilling the contract would have to be so high that (in residential construction) contracts are normally left quite "loose," especially as regard completion date.[29]

In other words, the cost of shifting all risks of inferior quality and delayed completion to the contractor would be so great that normally such risks are shared. Sharing implies leaving a penumbra of vagueness about quality and completion date, and relying upon the contractor's concern for his reputation and, in extreme cases, fear of law suits to insure reasonable performance. But effectively to judge the quality of contractor performance requires more knowledge and time than many buyers possess. To compensate for this incapacity, they hire an architect to supervise the construction. This normally implies closer supervision of quality than the buyer can provide, shifting some of the risk of quality variation from buyer to contractor; consequently a contractor's price for completing a given set of plans normally is higher if there is to be architect supervision than otherwise.

Where uncertainty of input prices, worker productivity [30] or availability of inputs is appreciable, the contractor may wish to minimize his own pecuniary risk and seek payment on a cost plus basis. Whether the buyer will accept such a contract obviously depends upon the price of alternative arrangements (e.g., a fixed fee or target price arrangement); his appraisal of the moral hazard in letting such a contract and his ability to supervise or oversee the work.[31] When a contractor (or worker) is paid on an hourly basis, he will usually

[29] An additional reason for permitting contracts to be loose, relevant to residential construction though not to medical care, is that the buyer often wants freedom to change specifications.

[30] Which varies with working conditions, the difficulty of the job, etc.

[31] The issue of deciding upon the form of contract also arises in the procurement of weapons. Most of the questions discussed in this section in relation to the purchase of medical care (e.g., evaluation and control of quality; speed of delivery; incentive to perform well; conflict of interest between buyer and producer) have analogues in weapons acquisition. For a detailed discussion of these problems in the context of the weapons industry see M. J. Peck and F. M. Scherer, "The Weapons Acquisition Process," Cambridge, Mass., 1962, especially Part III.

maximize his own satisfaction by working at a slower pace than that which would minimize cost per job at given quality. If his rate of pay is above the market, or he fears a shortage of demand at the going rate, pecuniary are added to nonpecuniary incentives to work slowly. Conversely, if he has mistakenly accepted a rate below the market and demand has unexpectedly increased, he has an incentive to speed completion at the expense of quality.

It is not usually feasible to vary payment to the contractor with quality of performance. But among lawyers, fees contingent on awards are very common in damage suits, and paying commissions as a percentage of sales is standard practice in a wide variety of selling situations. While the idea of relating the size of the doctor's fee to the success of the treatment is not unthinkable,[32] it is surely uncommon and not likely to gain favor in the near future.

But to say this is far from conceding that quality of treatment is independent of the method of payment, or that improved methods of payment might not help improve both quality and quantity of treatment. Contrary to the spirit of some recent discussion, the relation of pecuniary incentives to quality of care is not settled by noting the physician's dedication to his calling.[33]

Granted that the great majority of physicians are devoted to the welfare of their patients, granted even further that where the voice of ethics grows faint it is powerfully reinforced by fear of malpractice suits; these considerations do not distinguish physicians from lawyers, certified public accountants, architects and others whose integrity and presumed disinterest serve as the basis for investing large sums of money. Yet few clients of these other professionals would agree that pecuniary incentive was irrelevant to the quality of their performance.

Let us consider a few of the situations in which method of payment may plausibly be supposed to affect quality of medical care: (1) *Fee splitting.* Over the years there has been a steady murmur of complaint, within and without the medical profession, about unnecessary sur-

[32] As in the possibly apocryphal story of the Chinese who allegedly paid their physicians only when they were well.

[33] Arrow, *op. cit.*, especially pp. 949–951 and 965–966. Arrow expresses indebtedness (p. 949, n. 15) to the work of Talcott Parsons, *The Social System,* Glencoe, Illinois, 1951, who adopts a similar position on this matter.

gery and its relation to the need of surgeons for more business.[34] In at least one famous example, hospital utilization rates for patients in a comprehensive prepayment plan were well below those for a comparable group in the same community covered by a fee-for-service plan; [35] the major source of difference between the two groups was three diagnoses (tonsillectomy, appendectomy and hemorrhoids) where there is a high percentage of cases in which surgery is optional. The widespread nature of fee-splitting between the surgeon and the referring physician is indicated by the squabble between the American College of Surgeons on the one side, and the American Medical Association and the American Academy of General Practice on the other, as to whether surgeons might be permitted to hire the referring physician as an assistant and pay him a fee.[36] It is often alleged that fee-

[34] For example, Eli Ginzberg, "Physician Shortage Reconsidered," *The New England Journal of Medicine,* July 14, 1966, p. 86, writes:

> Here are some of the statements made in passing at the recent conference that went more or less unchallenged:
>
> Many general surgeons are not very busy and therefore a great amount of unnecessary surgery is performed. The situation is particularly shocking when hysterectomies are considered. We were told that many women undergo mastectomies when a less radical procedure would do. Many thyroidectomies are performed when psychotherapy would be preferred.
>
> There is substantial overdoctoring for a host of diseases, including in particular infections of the upper respiratory tract.

C. B. Esselstyn, M.D., "Principles of Physician Remuneration," in National Conference on Labor Health Service, *Proceedings,* 1958, p. 126, says:

> As the very carefully documented experience of the UMWA Welfare and Retirement Fund has so adequately proven, the medical profession today does not have sufficient maturity to resist the temptation of unnecessary procedures so inherent in fee-for-service. Furthermore, the greatest rewards under this method of payment go not necessarily to the person who is practicing the best quality, but rather the greatest quantity of medical care. And, as long as medicine is practiced on a fee-for-service basis, preventive medicine will never be the driving force of any doctor or group of doctors, and public health services will always be resisted for fear of encroachment.

[35] P. M. Densen, E. Balamuth and S. Shapiro, *Prepaid Medical Care and Hospital Utilization,* Hospital Monograph Series No. 3, American Hospital Association, Chicago, 1958; especially pp. 28–34.

[36] The statements and counterstatements may be found in *The New York Times* of October 5, p. 39; October 9, p. 34; October 17, p. 30; October 18, p. 42 and November 30, p. 27 (all dates refer to 1961). The battle involved a whole set of interrelated issues concerning the qualification of nonspecialists to perform surgical procedures, how much surgery is unnecessary, and what constitutes "fee-splitting," a name abhorred by all concerned. For a good summary of the controversy, see Lawrence Galton, "The Doctors Debate Fee-Splitting," *The New York Times Magazine,* March 4, 1962, pp. 19 *et seq.*

splitting is part of a make-work pattern involving referrals for unneeded surgery.[37]

(2) A pervasive effect on the allocation of physician time may be exerted by the fee for service (vis-a-vis prepayment) method of compensation. Under fee for service no reward is given for time spent in continuing study in order to keep up with new developments. Consequently, one would expect that physicians practicing under fee for service would spend a smaller fraction of their time in studying, and a larger fraction in income earning patient care. There is reason to believe that general practitioners spend less time in reading and updating their education than they would "like," or than expert opinion thinks they should.[38] Many of the prominent medical groups (who pay physicians a salary plus a share in profits) insist that their members allocate a definite amount of time, as a minimum, to continuing medical education. This evidence is supportive of the suggested hypothesis, but obviously it is not conclusive.

(3) A further implication concerning the effect of method of compensation on the allocation of physician time relates to the time spent on particular procedures, or the willingness to undertake procedures which are relatively undercompensated. Price theory suggests that substitution of payment by fixed fee [39] for fee for service (where the service unit is roughly proportional to time spent) will tend to discourage use of time-intensive procedures. And there is evidence that this has happened; Clute reports a number of complaints of Canadian physicians that some insurance companies refuse to pay adequate compensation for patient histories.[40] As a result, "These men pointed

[37] In addition to Ginzberg's statement (n. 35), see V. G. Damon and Isabella Taves, "Fee Splitting, Knife Happy Surgeons and Unnecessary Doctors," *Look Magazine,* June 19, 1962, pp. 86 *et seq.*

[38] For example, see Clute, *op. cit.,* pp. 463–479, or Peterson *et al., op. cit.,* pp. 90–91.

[39] By fixed fee, I mean any method of compensation independent of the time the physician spends or the amount or quality of service he renders. Obviously, many schemes will meet this criterion; payment by salary or by a capitation system (as in the British National Health Service) are two examples of what I loosely term "fixed fee" methods of payment. M. I. Roemer, "On Paying the Doctor and the Implications of Different Methods," *Journal of Health and Human Behavior,* Spring 1962, pp. 4–14, distinguishes three methods of payment: fee for service, capitation and salary. In effect I put both of the latter under fixed fee. James Hogarth, *The Payment of the Physician,* New York, 1963, gives a detailed description of methods used to pay physicians in various European countries.

[40] Clute, *op. cit.,* pp. 198–200.

out that if a doctor knew from previous experience that he would not be paid for the extra time that he took to investigate a case thoroughly, he would tend, through sheer economic necessity, to do the less time-consuming and less thorough work for which he was being paid." The implications of this, as far as quality of practice is concerned, are serious.[41] In Britain, the difference in treatment of private patients and of patients under the National Health Service is well-known and is what economic considerations would lead one to expect.[42]

I am not suggesting that the preceding scraps of evidence serve as the basis for any firm conclusions regarding the comparative effects of alternative systems of remuneration upon the pattern of medical practice. They are intended merely to indicate that there is some empirical evidence to support the implication of price theory that the physician's allocation of his time among alternative uses will respond in a conventional manner to variations in their relative rates of compensation per hour spent.[43]

The Market Mechanism and Quality of Care: A Digression

Like the general public most, though not all, economists would agree that the market mechanism cannot be trusted to provide suitable quantity and quality of medical care. The obvious reasons are (1) consumer ignorance of quality, fostered by professional "ethics" against advertising and public criticism of other doctors, and (2) restrictive practices that impede entry. Presumably, the restrictive practices could be eliminated, but relative consumer ignorance cannot help but increase with growth in the stock of knowledge. These reasons are relevant and may be adequate as a rationale for this belief.

However, there is another factor that enters into the prevailing

[41] *Ibid.*, p. 470.

[42] Paul Ferris, *The Doctors,* London, 1967, pp. 33–52 and 159–163.

[43] Roemer, *op. cit.*, gives emphatic support to the view taken here of the incentive effects of the fee for service method of payment.

An interesting example of the consequences of a lack of pecuniary incentive to perform well (combined with inadequate supervision) is described in a careful case study of the behavior of hospital nurses when the number per shift was increased in the hope of improving specific aspects of quality of patient care. (M. K. Aydelotte and M. E. Tener, *An Investigation of the Relation Between Nursing Activity and Patient Welfare,* Iowa City, 1960, especially Chapter IX.) The principal effect of the increased staffing was found to be mainly in increased leisure on the job for the nurses, and there was little increase in time used for patient care.

attitude on this question which is frequently overlooked. This is the difference in attitude commonly taken toward the quality of medical care we all should get—the best—and that which we are willing effectively to demand as individuals. The cost of obtaining superior medical care is not primarily—or perhaps not at all—a matter of pecuniary expense, but the time, trouble and great difficulty of identifying it. Moreover, high quality care is likely to be relatively time consuming, for patients as well as doctors; [44] careful examinations also create patient anxieties. As a result, many people who should be able to judge physician quality may use inferior but "convenient" physicians, others don't use physician services often enough. Many people of my acquaintance are quite casual about the selection of doctors until they, or someone in their family, become "really sick" whereupon they take stock and not infrequently change physicians.

Such behavior may reflect adversely on the judgment of the individuals involved. But this is not beyond dispute; granted the importance of early diagnosis and treatment of certain diseases, the increase in the expected speed of discovery as a result of using a higher quality physician, combined with the difficulty of identifying him as such, may not be great enough to "justify" the time and trouble of a busy man.[45] Moreover the psychic, not to say psychosomatic, costs of carefully scrutinizing physician performance may not be negligible.

The point of this digression is not to discourage concern with quality of medical care nor to advocate unquestioned reliance upon consumer choice in directing the allocation of physician time. Rather it is to stress the uncertainty of the strength of the public's revealed preference for superior medical care. Clearly most people will insist on having access to good—even "the best"—medical care once it is identified as such. But many of these same people continue to smoke cigarettes, and so do superior doctors. The evidence that links cigarette smoking to lung cancer, heart and other diseases is at least as strong as that relating physician quality to any or all aspects of health.

[44] The importance of consumer time as an element in the cost of consuming a particular item is stressed by G. S. Becker, "A Theory of the Allocation of Time," *Economic Journal*, September 1965, pp. 493–517.

[45] A spectacular illustration of this point is afforded by the failure of physicians to detect early cancer in themselves. See Robert Sutherland, *Cancer: the Significance of Delay*, London, 1960, pp. 196–202.

The already weak link between physician quality and health is further attenuated, and greatly, by the difficulty of identifying physician superiority.[46] Consequently, it is neither surprising nor necessarily irrational to prefer a pleasant manner or some combination of lower fees and more time per contact to brusque efficiency hopefully combined with superior knowledge. On the other hand, I would not be prepared to bet that there would not be a substantial shift of demand (at current fees) if the rankings made in a study of the Peterson-Clute type were made public. Moreover, within limits (stated below) I would favor improving quality of care whatever the state of consumer preferences.

In a nutshell, a very important obstacle that the market must overcome in delivering an "adequate" supply of high quality medical care is the scant premium in time and trouble the individual consumer is prepared to pay for superior quality. This is not at all incompatible with a strong consensus that government—or somebody—ought to make the effort to insure that all dispensers of medical care be of high quality. Indeed, I suspect that much of the demand for licensure of medics springs from the (mistaken) belief that this insures high quality care regardless of how a (licensed) physician is selected.[47] But so long as prospective patients act as though they didn't care much about physician quality, the market cannot be expected to guarantee that physicians will meet the "extra-market" quality standards of public health officials, medical school professors and concerned citizens; hence the role for presumably expert intermediaries, such as health plans, in improving quality of care.

Productivity and Quality of Care

It is frequently argued that group practice is both more efficient and provides a higher standard of care than solo practice.[48] If so, the

[46] The difficulty of appraising the quality of a physician can be readily appreciated by anyone even casually perusing evaluative studies such as those of Peterson *et al.*, *op. cit.*, or Clute, *op. cit.* And these very careful studies did not attempt to estimate the differential effect on patient health of differential physician quality.

[47] Clute, *op. cit.*, Chapter 25, gives an excellent discussion of this point and offers proposals for revamping the entire system of licensure.

[48] For example see Fein, *op. cit.*, pp. 94–104, or H. M. and A. R. Somers, *Doctors, Patients, and Health Insurance*, Brookings Institution, Washington, 1961, pp. 487–488. Klarman, *op. cit.*, pp. 126–129 presents a balanced appraisal of this contention.

substitution of group for solo practice will tend to increase measured productivity. But granting for the moment that group practice is more efficient (in terms of input-output relations), it is by no means clear that quality of care is superior. The allegations of Fein, Somers and Somers and others concerning the superior quality of care under group practice refer only to certain technical aspects of care. Weinerman states that in the few studies made thus far

> more personalized care is ascribed to the independent practitioners and better technical standards to the group physicians. . . . The meager evidence to date suggests that independent and general practice is more acceptable from the patient's point of view, while specialty and group practice provide an objectively more rational and efficient standard of professional service.[49]

Though the evidence to support it may be meager, this statement is not implausible. If it (the statement) should be true, it follows that although the secular drift toward specialization and group practice may have improved the quality of care in a technical sense, it may have worsened it in terms of the "incidental" satisfaction furnished (some) patients. Many, perhaps most, people would insist that the gain far outweighs the loss. That is, many people believe it is the principal business of doctors to heal those "really" sick (i.e., having specific organic malfunctions) and that what they do for hypochondriacs, or individuals with vague complaints of which they are unlikely to die, is decidedly less important. While I do not necessarily dissent from this value judgment, I should like to point out some of its implications.

The outlook on health problems that underlies this value judgment rigidly dichotomizes people into the (organically) ill and the healthy. The sick may be further divided into those who can be helped and those who can't. In this view of the world, the doctor is best fulfilling his function when he is tending the curably ill. There is considerable evidence that the utility function of the typical doctor reflects this value judgment. Doctors have been found to express an overwhelming preference for dealing with patients with specific organic complaints; in the medical sub-culture the "crock" is abhorred from

[49] E. R. Weinerman, "Research into the Organization of Medical Practice," *Milbank Memorial Fund Quarterly*, October 1966, pp. 117–118.

student days on.[50] Thus physicians tend to give priority to those who "need" them, and whom they know how to help.

But what if, as many believe, there is no simple dichotomy between organic and inorganic illness? It may well be that many complaints, categorized as psychogenic, have an organic basis that is difficult to ascertain. Because it is difficult to diagnose these complaints, and because often they are psychosomatic, the physician is uncertain as to whether there is a course of treatment that will benefit the patient and believes that even if there were, he would have little chance of finding it. In other words, in deciding how to allocate his time among his patients, the physician is confronted with a distribution of probabilities of benefiting patients, each probability conditional upon the initial diagnosis. The highest probabilities are for diagnoses in which he is confident and for which there is a generally accepted method of treatment; doubt either as to diagnosis or to treatment reduces the probability of a successful outcome.

The marginal utility, to the physician, of using his time to help a given patient depends not only upon the compound probability of being able to make a correct diagnosis and find a course of treatment that will improve the patient, but also upon the degree to which he thinks the patient will benefit, upon the importance he attaches to the contribution he believes he can make to the patient's welfare, and upon the fee he earns. A physician who maximized pecuniary income would be concerned only with the last item, and would allocate his time so that an extra minute spent with any patient would yield the same income. This is a possible model of behavior, but I doubt that many doctors follow it.

One reason is that the growing importance of medical care insurance has made it increasingly difficult to devote more time to one patient than to another with the same illness, and charge correspondingly more for it.[51] Another reason is that most doctors have always

[50] See W. Martin, "Preferences for Types of Patients," in R. K. Merton, *et al.*, *The Student-Physician*, Cambridge, Mass., 1957, pp. 189–197; S. W. Bloom, *The Doctor and His Patient*, New York, 1963, p. 36 (and the references in p. 51, n. 9); H. S. Becker *et al.*, *Boys in White*, Chicago, 1961, pp. 316–329; O. L. Peterson *et al.*, *op. cit.*, p. 41.

[51] Charging a higher total fee for spending a larger amount of time is not discriminating; indeed, under a fee for service method of payment, *failing* to charge more for more time spent would be discriminatory. However, insurance companies

sought the nonpecuniary satisfactions of professional esteem and patient gratitude, as well those of money income.[52] These satisfactions are mainly those of healing, and of contributing to medical knowledge; in this the physician's utility function is quite similar to that of an economist or of almost any other professional. In pursuit of nonpecuniary satisfactions, physicians tend to seek either those cases where they are confident their skills will prove fruitful (i.e., where they expect to derive the satisfactions of success), or those which present interesting diagnostic or therapeutic challenges.[53] Also, important people whom the doctor wishes to cultivate for social reasons, because of the indirect effect on his practice or for other reasons, get a disproportionate share of his time.[54]

But the ordinary middle-class person with a variety of vague middle-

and health plans may refuse (see Clute, *op. cit.*, pp. 198–200) to pay more than a prescribed maximum for treating a specified condition.

It is generally agreed that physicians discriminate in setting fees, charging the rich more. The fact that some of their service is rendered free (for charity) makes the allegation of price discrimination highly persuasive. However if, as seems plausible, they spend more time with richer patients and give them prompter appointments (i.e., cost them less in waiting time), then failing to charge them somewhat more than poorer ones would be discriminatory. This does not destroy the arguments of Kessel (R. A. Kessel, "Price Discrimination in Medicine," *Journal of Law and Economics,* October 1958, pp. 20–53), but it does reduce their force.

[52] In one sense this statement is obvious; there is an undeniable validity to the Arrow-Parsons contention that the role of physician requires that those who perform it consider the interest of the patient as well as—or ahead of—their own. However, this image has been tarnished by the wide variety of money-making expedients to which physicians have resorted, and the hypocrisy that has surrounded the activities of their professional organizations. Citations in support of the statement in the text could fill a volume, but not a few could also be found that would support dissent. Rather than offer a long and inevitably indecisive argument, I shall let the statement in the text stand unsupported.

[53] The importance of the interesting case, both to medical students and practicing physicians, is indicated in *Boys in White, op. cit.* The concern with "interesting cases" is an aspect of the increasingly scientific preoccupation of the medical community; it reflects "disease orientation" rather than "patient orientation."

[54] In extreme cases, such as the royal family in Britain, the amount of attention paid by physicians is notorious.

> "The late Lord Evans is said to have encouraged the privileged public to treat consultants as GPs, with his readiness to go to Buckingham Palace for trivial matters that could have been left to other and lesser royal doctors. 'He was a hell of a nice guy, but if anyone sneezed at Buck House he used to go there,' said another doctor disapprovingly." Ferris, *op. cit.*, p. 36.

In the United States, the President has a resident physician in the White House. The pride of the physicians of movie stars and the like strongly suggests that these luminaries do not lack medical attention.

aged complaints is likely to find doctors curt, expensive (per minute of contact), difficult to see, and not helpful. The physician's defense is that he can't do much for them, and probably there is nothing organically wrong with them anyway. Granted both contentions, there is still some chance that if a physician took the time and trouble thoroughly to investigate these vague complaints they could be remedied. And I suspect that many people now denied this type of treatment, would be willing to pay a "market price" to obtain it.[55] However, they are thwarted by physician unwillingness to embark upon such programs of uncertain therapy.

Such (alleged) reluctance by physicians does not indicate irrationality. The alternative, billing a patient for the full cost of a physician's time, would often result in unusually high charges which patients might be reluctant to pay if—as is very probable—the treatment did no good. Complaints about physicians who keep patients returning for treatment without benefit are legion. It is easy, and sometimes correct, to blame physician greed for such occurrences; but it is often hard to know whether it was the physician or the patient who urged the course of treatment, and what was the understanding on which the treatment commenced.[56]

Elaborate advance contracts to insure payment would surely put the physician in an unfavorable light with his professional peers as well as with laymen. Similarly, incentive type contracts, with a great premium placed upon success, would run grave risk of being considered unethical,[57] as well as encountering serious problems of proving

[55] The difficulty of getting a physician to make a housecall is indicative of the situation. High though the cost of housecalls may be in most locations, they do not appear to have gone high enough, relative to the prices of office calls or hospital visits, to induce physicians willingly to make them.

[56] See R. H. Blum, *The Management of the Doctor-Patient Relationship,* New York, 1960, especially pp. 132–138.

[57] The medical world view is nonprobabilistic and strongly biased in favor of avoiding errors of commission ("do no harm") even at the cost of increasing the probability of errors of omission. (See T. J. Scheff, "Preferred Errors in Diagnosis," *Medical Care,* August 1964, pp. 166–172.) Moreover, the physician's role as expounded by Parsons, *op. cit.,* is probably irreconcilable with a game theoretic approach to therapy which was frankly avowed to the patient. And the physician's superior knowledge would make basing the size of fee on the outcome of the treatment most unfair to the patient. But, fair or not, unusually high fees (reflecting improbable cures) would probably be rejected by the courts, and frowned on by medical societies.

success. Hence it is prudent to avoid this whole set of problems by sticking to conventional courses of therapy and conventional bills.

The consequence of this is to ration time so as to discourage "crocks." This is normally done by making them wait long periods for appointments,[58] and being brief when dealing with them. This has the effect of raising the marginal cost, to an undesirable patient, of obtaining a given amount of a physician's time without raising the pecuniary return to the physician. This apparent paradox reflects the fact that the process of producing many services, including medical, involves the utilization of a variety of unpaid inputs. Of particular importance is the time spent in travelling to the physician's office and in waiting there.[59]

Waiting for an appointment also involves some costs not usually appreciated. Delivery next week is not as good as immediate delivery; in a world where every good and service could be bought in a market, a commodity delivered next week would be different from one delivered today and would be bought in a different (though related) market. Delayed appointments with a doctor are billed at the same fee as prompt ones; the queueing procedure is usually "first come, first served" for available dates, with "emergencies" being taken out of turn, causing increased waiting time to those already scheduled.[60]

The effect of this time rationing system, on the measurement of changes in output and productivity in the medical care industry is as follows: secular increases in productivity in part reflect a shift in the locus of treatment from the home to office and hospital. This shift has made it possible to save the doctor much of the time formerly spent in travelling from house to house; [61] the same shift has also

[58] I.e., make them wait a long period to get an appointment and sometimes make them spend a long interval in the waiting room also.

[59] This has already been remarked in n. 44. Also see V. R. Fuchs, "The Contribution of Health Services to the American Economy," *Milbank Memorial Quarterly*, October 1966, pp. 73–74.

[60] Of course, the doctor has considerable latitude in deciding what is to be considered an emergency.

[61] Somers and Somers, *op. cit.*, pp. 48–49 present data suggesting that home visits declined from 40 per cent of all nonhospital visits in 1928 to 8 per cent in 1957. Mrs. A. A. Scitovsky (*Changes in the Costs of Treatment of Selected Illnesses, 1951/52 to 1964/65*, Palo Alto, California, May 1967, Table 8, p. 68) reports that in 1951–52, 34.7 per cent of all physician visits in treating otitis media (in children) were made at home; in 1964–65, the corresponding percentage was 3.4.,

made it possible for the doctor to be supplemented by a battery of assistants (nurses, laboratory technicians, secretaries, etc.) who further enhance his productivity. But this gain is at least partially offset by the increase in time spent by patients in travelling to and waiting for the doctor; this increased input should be considered in measuring productivity change.

The increase in productivity in medical care has been indissolubly connected with a great improvement in the nature of the product and, because of a combination of rising income, education and medical care insurance, a vast increase in demand. For reasons already indicated this has involved a relative shift in supply of physician time to those illnesses that are capable of effective treatment and away from those that are not. This has had (see below) the effect of reducing the availability of medical care to higher income groups relative to lower, which is tantamount to a redistribution of real income. In other words, the increase in measured productivity has been partially due to a substitution of poor-man's medicine for rich-man's. Allowing for this will inevitably reduce estimates of the productivity increase that would have occurred with an unchanged, homogenous output.

PROLEGOMENA TO A MODEL OF THE MEDICAL CARE SECTOR AND ITS POLICY IMPLICATIONS

It is a truism, or should be, that to measure productivity change it is necessary to have an explicit model of the economy or economic sector under consideration. It is convenient to define productivity changes in terms of shifts in a production function, but to speak of a sector's (firm's, economy's) production function is to assume that the problems of specifying its form and operationally defining its variables have been solved. But these problems are never simple, and that is especially the case in this sector. As its title indicates, this section does not attempt to offer a model of the medical care sector. It attempts only to discuss certain peculiarities of medical care production that must be considered in constructing such a model.

The Entrepreneur and His Objectives

The two hardest questions in the theory of production are "who is in charge here?" and "what is he trying to do?" These questions are

so difficult, and the answers appear to vary so much from one situation to another, that textbook discussions almost invariably bypass them and hasten on to discuss the constrained optimization of a specified maximand. Here, let us attempt to face the harder questions.

Identifying the man in charge of dispensing medical care seems easy. By law, custom and common belief, medical care is what the doctor orders.[62] Reality is not that simple: the relation of physician to surgeon; of both to hospitals, medical plans, and insurance companies greatly complicates the situation. Nevertheless, the actual position of the doctor and expository convenience both suggest that the doctor be considered as the entrepreneur. This specification refers to the United States at the present time and the past; but it is not the only conceivable arrangement, and the situation is subject to change.[63]

Assuming the doctor to be entrepreneur of a firm producing medical care,[64] what are his objectives? To circumvent methodological dispute, assume him to be a utility maximizer, but compelled to operate under special constraints. (If one wished to posit that he was a satisficer with respect to money income, the subsequent discussion would not be altered.) One particular and very important constraint upon the physician's actions is the heavy penalty imposed for certain types of unconventional behavior, without adequate counterbalancing rewards for unusual achievement that might result from such behavior. This introduces a marked aversion to particular types of uncertainty. For example, failure of an unorthodox method of treatment may bring malpractice suits, the hostility of the local medical society, or both, which can severely injure a physician by denying him referrals, access to hospitals, and even loss of license to practice. Such loss is very similar to "gambler's ruin" in a fair bet.[65] At the very least, failure of a long course of treatment is likely to make it difficult to collect a bill for the value of physician time spent. On the other hand, success

[62] For example, in the American Medical Association's study of the cost of medical care the physician is identified as a "contractor" of goods and services for the patient. *Report of the Commission on the Cost of Medical Care,* Chicago, 1964, Volume I, p. 16.

[63] See, for example, the recommendations of Clute, *op. cit.*, Ch. 25.

[64] Implying that hospitals, consulting specialists, etc., are all hired or purchased inputs.

[65] I.e., loss of his stake, so that he is unable to continue playing.

against long odds can gain the *practicing* physician very little; [66] because he cannot advertise his success, he cannot profit through higher fees or more patients, nor can he block the swarm of imitators. As a result, the practicing physician has a strong tendency to encourage patients with standard organic ailments for which there is a recognized course of procedure, preferably one that brings symptomatic relief.[67] This brings him kudos, both from patients and other physicians, adding to his money income and to his nonpecuniary satisfaction. Conversely, he will discourage other types of patients.[68]

Utility maximization by the physician leads to rather special behavior in the field of price setting. Although doctors are in many respects a classic example of a guild, they have not been greatly concerned with maintaining minimum prices. This is due partly to the universal acceptance of price discrimination, and partly to the fact that continuing to serve in the face of nonpayment or partial payment is an effective form of de facto price cutting. Instead of focussing attention on price behavior, the medical profession has stressed the importance of maintaining high professional standards which has involved limitation of numbers.

By its system of induction through medical schools, hospital appointments, etc., the profession inculcates an image of the successful doctor whom neophytes strive to emulate. These paradigms are not money income maximizers, though their incomes are notoriously high. Given the opportunities for price discrimination inherent in medical practice, money income maximization would imply an attempt to eliminate each patient's consumer surplus from being treated by the

[66] This is not the case in a *research* situation. However, the ordinary practicing physician is not engaged in research and those who are so engaged are interested only in patients with ailments germane to their research. The person who encounters a doctor with a research interest in his complaint is indeed fortunate, and gets a kind of treatment normally unobtainable. Of course, such a patient must bear with the inconveniences that typically attend a research situation.

[67] Organic complaints are preferred because psychosomatic patients are often psychologically disturbed, and suit-prone.

[68] The conservative bias of the physician, both in selecting patients and in treating them, strongly discourages attempts to increase money income by deliberately giving substandard treatment, although there are spectacular examples to the contrary. Detectably substandard performance incurs risk of collegial disapproval with consequences already noted.

doctor in question rather than by any other physician.[69] But this is not what is attempted.

One very important reason why it is not attempted is that hospitals and medical schools are eleemosynary institutions which draw their support from the very same social groups whose members pay the highest fees. The leading medical practitioners who treat the wealthiest patients are symbiotically associated with these institutions, both drawing and lending prestige through the affiliation. Consequently, they must leave their patient-patrons feeling pleased with their treatment, personal as well as medical. This institutional necessity has much to do with the "collectivity-orientation" emphasized by Parsons.

A second reason why physicians do not maximize money income is that the professionally approved doctor-patient relationship is one whose terms are dictated by the doctor. Ideally, the physician is supposed to treat the patient in accordance with the precepts of his calling, irrespective of what the patient wants; the dissatisfied patient is to be invited to go elsewhere. The patient is to feel privileged that his physician will treat him and is to reward the physician not only in money but also by following instructions. This relationship requires that the physician be virtually indifferent as to whether a patient leaves him.[70] Obviously this is most likely to happen when the physician has more patients than he wants and is therefore inconsistent with maximizing the amount of money income he obtains from each one and hence in total. Unless the physician is in this position vis-à-vis his patients he will be tempted to perform operations or prescribe drugs that patients demand, even though in his judgment the treatment would do harm.[71]

The posture of "dominating the patient through economic independence" is not possible for all doctors. Beginners, those inept at attracting patients, those in locations of relative excess supply and the greedy everywhere are not able to follow the light of professional con-

[69] This does not imply that the physician could succeed in this goal, even if he sought it. If he sought this objective, the physician and each of his patients would be in the position of bilateral monopolists.

[70] If the physician charged all patients the same price for a given service, this condition would imply that he charged less than the market equilibrium price in order to ration services among his patients.

[71] On this point, see the remarks of Damon and Taves, *op. cit.*

science single-mindedly. Normally, however, as a physician becomes established his case load and income rise until he can comfortably assume the approved posture of economic independence. Indeed, pride or group pressure often cause assumption of this posture well before it is financially appropriate.[72]

A final word on income maximization. Professional ethics refers to the relation between a doctor and his own patients or, at most, to his obligation to those whom he encounters in an emergency. It does not require him to accept particular patients or to locate himself where he might be needed. In a forthcoming paper, two of my colleagues and I find substantial evidence for the proposition that the *locational decisions* of physicians are compatible with a desire to maximize money income.[73] However, this is not incompatible with pursuit of other objectives in fee setting and allocation of work time.

Method of Payment and Its Consequences

The conventional theory of production assumes that the productive efficiency of an input unit is independent of its rate of reward. The performance of both human and nonhuman inputs is assumed to be completely insensitive to the effect of pecuniary incentives. The impact of incentives falls solely upon the entrepreneur who is assumed to be operating upon the production function but whose services are not among its arguments. The effect of incentive payments on the productivity of employed workers remains largely unexplored in economic theory though there have been many incidental observations on the subject.

The key role of the physician in the production of medical care would suggest that his method of compensation would affect his output. And, as we have seen, it does; but it would do so to an even greater degree if physicians were money income maximizers. The desire to achieve professional esteem leads him to charge less than the "traffic will bear." This enables the physician, in both fee for service

[72] Accepting a salaried post (e.g., with a group of physicians) while building contacts with patients and colleagues and accumulating some capital often enables the beginning doctor to reconcile the demands of professional rectitude with economic necessity.

[73] L. Benham, A. Maurizi and M. W. Reder, "Location and Migration of Medics: Physicians and Dentists," *Review of Economics and Statistics,* August 1968..

and salaried situations, to resist patient demands and to give treatment in accordance with conventional medical standards.

However, since fee for service patients generally pay more per "incident," they expect and receive more of physician time, and can insist upon being treated by their own doctor. This is in accord with accepted medical practice and were it not the case few patients would voluntarily choose this more expensive method of care. But the amount of physician time the patient can demand, even when he is willing to pay standard hourly rates, is rationed by the physician who also controls the kind of treatment offered.

It may be that general surgery provides an important exception to this picture. As already indicated it is widely believed that shortage of demand at prevailing fee schedules has frequently led to medically inadvisable operations. This belief is consonant with the overwhelming opinion of surgeons that there is an excess supply of general surgeons, but that such excess does not exist in most of the various specialized branches of surgery.[74] The belief is further supported by the frequent allegations that surgeons split fees with referring physicians, and by the complaints of physicians that per hour of time surgeons are greatly overpaid relative to themselves.[75]

In brief there is a variety of reports concerning general surgery that suggest the existence of the type of situation that would arise if a service were priced higher than the level at which demand equalled supply: i.e., attempt to stimulate demand for the service (by encouraging unnecessary operations), price-cutting (fee-splitting) and attempts to exclude competitors (nonspecialists) from the market. Superficially at least, it would seem that the market situation in surgery was quite different from that in the other branches of medicine.

The contrast between the (allegedly) loose market in surgery and

[74] In 1962, the American College of Surgeons sent a mailed questionnaire to 20,000 of its fellows of whom 65 per cent replied. In substance the question asked was whether the respondent thought there were in his community "too many," "just enough," or "not enough" of the following types of surgeons: General Surgeons; Obstetricians-Gynecologists; Orthopedists; Urologists; Ophthamologists. An overwhelming majority of both General Surgeons and other surgeons thought there were too many General Surgeons; but for all other fields (except Obstetrics) the prevailing view was that there were too few rather than too many practitioners. *Bulletin of American College of Surgeons,* November–December 1963, pp. 354–355.

[75] See Clute, *op. cit.*, pp. 205–210.

the tight one in most other branches of medicine is quite in keeping with the behavior pattern underlying the allocation of hospital beds mentioned earlier.

The Eleemosynary Character of Medical Care

A predominant characteristic of the medical care industry is its eleemosynary nature. This characteristic affects not only physicians but auxiliary personnel as well. Nonprofessional workers (orderlies, nurses' aides, etc.) have always been among the lowest paid workers in the community. This is partly because of the unskilled nature of the work, but it is also a reflection of the ability of eleemosynary institutions to escape the minimum wage laws and conventions that normally set floors under the hourly wage rates of this part of the labor force.

Yet another factor working to hold down labor costs in this sector is the large amount of unpaid labor it can draw upon. This is related to the great ability of this sector to lower the quality of the nonmedical aspects of hospital and office care. If nurses are too few, patients can be allowed to wait longer; if they are incompetent, patients may be inconvenienced but, hopefully, not injured medically. To some extent, deficiencies of auxiliary personnel can be offset by greater use of unpaid assistance, especially by members of patients' families. But in any event, supplicants dare not be complainers. This tradition is dying under the spread of hospital insurance, but slowly and hard.

The eleemosynary aspects of providing medical care affect demand as well as supply. Medical care has never been considered as an object of choice which one could consume as means and taste permitted. Until this century it was something that ought to be provided by charity, and more recently a service to which a human is entitled by right. This right has been enforced through public provision of medical care and through semicompulsory medical care insurance. Both government and insurance companies strongly resist increases in the price of medical services because of the need to hold down taxes or monthly payments. Since professional standards and public surveillance both limit deterioration in the quality of medical care per se, the pressure of cost increases is deflected, at least partially, from raising product price to lowering the amenity aspects of quality.

Some Conjectures on the Recent Economic History of Medical Care

Though I cannot prove it, I suspect that prior to World War II rather few doctors were in the position of being independent of a credit worthy patient. Many (most?) patients were in the position of being in arrears or paying "what they were able," and toward them the doctor could maintain the posture of independent benefactor if he wished. The pre-World War II position of the doctor was also greatly affected by the long secular adjustment to the rise in professional standards following the Flexner report: this adjustment was reflected in a gradual decline in the over-all physician-population ratio as well as in a decided relative improvement in the qualifications of the lower part of the profession.[76] But in the meanwhile (circa 1915–1945), the large stock of previously licensed (largely unspecialized) physicians continued to render time consuming if ineffectual service to middle-class citizens willing to pay the standard fees. Others who could not pay full fees, were treated in a manner varying with the attitude of the physician.

Changes in this situation, occurring by the slow process of attrition, were greatly accelerated by World War II. The postwar physician was appreciably better trained, more specialized and more strongly oriented toward his colleagues, their hospitals and the standards of the profession than his predecessor. Because of improvements in apparatus, chemotherapy and drugs he was much better able to obtain visible results from treatment.

On the demand side, he was confronted by a clientele that was better educated and with far higher per capita incomes than their predecessors. Moreover the social reforms of the 1930's and the substantially full employment maintained since 1940 worked to increase family income most in that sector of the population where income elasticity of demand for medical care was greatest. Demand was further enhanced by the spread of health insurance and prepayment plans which served both to lower the marginal cost of care and to spread knowledge of its potentialities.

The percentage increase in demand for medical care at constant

[76] On this point see M. Friedman and S. Kuznets, *Income from Independent Professional Practice,* New York, NBER, 1945, pp. 8–21; also R. H. Shryock, *Medical Licensing in America, 1950–1965,* Baltimore, 1967, pp. 55–76.

prices was far greater than the percentage increase in number of physicians during the postwar period. In response to this, on-going changes in the pattern of practice were greatly accelerated; house calls were drastically reduced in favor of office and hospital visits; in office and hospital, assistants of all kinds both economized the doctor's time and minimized his personal contact with patients. The tendency toward depersonalization of the doctor-patient relation was further accelerated by the steadily increasing degree of specialization and group practice which replaced "the" doctor by a team, none of whom remembered the patient's name or face.

The increasing degree of specialization has facilitated growth of physician income. Services previously performed by an individual, but less specialized, physician have been increasingly delegated to specialists each of whom sends a bill. This increased division of labor, has often reflected the application of new techniques of treatment, which probably improved the quality of care. However, this same division of labor led to a more rapid increase in the cost of curing specific ailments than could have been inferred from the movement of the medical component of the CPI.[77]

The comparatively moderate rise in physician fees (as recorded in the CPI) since World War II has often been noticed.[78] In part, this (relative) stability is misleading because of improved collections, and because relatively more patients paid the standard fees reflected in the CPI.[79] But some of the same forces that improved collections and

[77] See A. A. Scitovsky, *op. cit.*

[78] For example, see J. W. Garbarino, "Prices and Productivity in the Medical Market," *Industrial and Labor Relations Review,* October 1959, pp. 6–10.

[79] A. A. Scitovsky, *op. cit.,* Klarman, *op. cit.,* pp. 151–153. Mrs. Scitovsky ("Changes in the Costs of Treatment," *op. cit.,* Table 7, p. 18) reports that at the Palo Alto Medical Clinic, in 1951–52, in treating otitis media, the average fee paid for an office visit was 68.5 per cent of the customary or standard fee and the average fee for a home visit was 88.8 per cent of the customary fee. In 1964–65, the average fee was 89.3 per cent of the customary fee for office visits and 102.8 per cent for home visits.

Mrs. Scitovsky also reports (*op. cit.,* Table 5, p. 46) that in 1951–52, the average fee for an appendectomy was 91 per cent of the customary fee; by 1964–65, the average fee for an appendectomy had risen to 100.5 per cent of the customary fee. In 1951–52 the average for a radical mastectomy was 89.8 per cent of the customary fee and in 1964–65, the corresponding percentage was 99.0 per cent (*op. cit.,* p. 117, Table 5). However, for maternity care in normal pregnancies the average fee was 100 per cent of the customary fee both in 1951–52 and in 1964–65 (*op. cit.,* p. 81, Table 4).

diminished the number of charity or quasi-charity patients have also inhibited increases in scheduled fees; i.e., the relative growth of third party payers (insurance companies and government agencies). Rather than engage in complicated negotiations with such entities, whose interest in physician goodwill is subordinate to that of minimizing payment of claims, increasing "productivity" has been an easier way of adjusting to market forces. The effect on physician incomes has been much like the increase in hourly earnings of piece workers that results from increases in output per man hour.

Productivity Gains and the Nature of Medical Care

Medical care has two aspects: that which deals with the organic ailments of a patient and that which provides the patient with various amenities, including physician time to explore various personal problems only uncertainly related to physical distress. Shifting physician time from the latter aspect to the former undoubtedly increases the number of patients treated and probably increases measured productivity per physician hour.[80] Such a shift may not reduce the health care available to anyone, or the benefits to the gainers may be considered to outweigh the harm done to the losers. As I have argued earlier, the implicit value judgment underlying the behavior of most physicians is either that treating organic ailments is more important than alternative professional uses of their time, or that they can accomplish more (per hour) by dealing with such ailments. Thus a sharp increase in the number of persons with specific organic ailments per physician available, as during World War II, may often lead to a parallel increase in physician productivity, as treatment is accelerated and crocks are discouraged more emphatically. This is what underlies Ginzberg's assertion that "the effective use of physician manpower depends in the first instance on a taut supply of physicians."

However, to say that there is a sufficient number of physicians to treat all serious organic disease that can be treated—if anyone said it—would not imply the absence of physician shortage. For if the attention of a physician is withheld from someone (with an ailment not

[80] This assumes that physicians do not charge patients in proportion to the time they give them, but charge less to those who use more time. If so, a shift to briefer appointments will increase measured productivity per physician hour.

having a specific organic diagnosis) willing to pay for it at its "market price," then in an important sense there is a shortage of supply. Let us grant the value judgment that the present stock of physicians should be allocated so as to maximize some function of "community health" into which each individual's health enters independently of his wealth. There may, nevertheless, be too few physicians to handle the nonurgent medical demands of reasonably well-to-do individuals at prices they would be willing to pay, and that would yield a sufficient return on the cost of physician training to induce a supply adequate to provide for these (presently unsatisfied) demands, without reducing the services anyone is presently receiving. What is suggested is the *possibility* that the institutional difficulty of charging market rates for treating patients with diagnostically difficult or psychosomatic illness requiring lengthy treatment of uncertain efficacy, may have lead to an underestimate of the demand for such treatment.[81]

Recognition of this possibility has been impeded by (justified) concern with the inadequate medical care available to low income groups. Concern over possible shortages of physicians and other inputs of medical care has almost always been allied with concern over the adequacy of care for lower income groups. But if the argument of this section is correct, progress in providing for the medical needs of the poor may have involved a reduction in the physician time available to other segments of the population. This reallocation of time is likely to be reflected as an increase in physician productivity; it may not affect adversely any indexes of physical health, but it will be associated with an altered state of patient-doctor relations that is one aspect of quality of care.

[81] The reader will note that I am suggesting the possibility of this particular market failure, not asserting that it has occurred. Looked at in a somewhat different way what I am suggesting is that given their high money incomes and the prevailing taste for agreeable patients, few physicians (other than psychiatrists) feel able to discriminate financially against unpleasant ones (crocks usually are, or soon become, unpleasant) sufficiently, to induce their willing acceptance except (perhaps) when demand is slack. Of course an undesirable patient cannot always be immediately identified as such, and since few physicians are willing summarily to dismiss a patient, they must discourage them gradually.

Unfortunately I can think of no practical way of testing this conjecture on observed market behavior, though it would not be difficult to develop experimental situations in which relevant evidence might be obtained.

DISCUSSION

HERBERT E. KLARMAN, The Johns Hopkins University

Professor Reder's paper represents a conceptual contribution to an important subject area in health economics—the definition of output and the measurement of change in productivity.[1] To the economist such matters are of prime intellectual interest; they are also central to the sensible consideration of such issues as the development of effective formulas for reimbursing hospitals, the soundness of actuarial projections under Medicare, future requirements for additional health services personnel, and the comparative costs of alternative methods of treating patients with mental illness.

PECULIAR CHARACTERISTICS OF INDUSTRY

The peculiar characteristics of medical care that require special consideration, Reder states, are consumer ignorance and the practice of price discrimination by producers.

Ignorance

Reder's discussion of the physician and his complex relationship with the patient properly extends far beyond an elaboration of consumer ignorance. One might add that sometimes the physician is ignorant, too, and must rely on empirical measures and on the placebo effect. The importance of forms of medical care organization, particularly the nonprofit hospital and its manifold relationship with the practicing physician, is noted. The prepaid group practice arrangement receives close to the almost standard "it stands to reason" treatment, in which its superiority is presumed in the absence of contravening evidence.

The discussion of the attitudes and role of the physician is interesting but, in my opinion, too accepting of the position of medical educators. The view that the physician's task is to treat and cure organic illness is too narrow. The physician can do other useful things, includ-

[1] Herman M. Somers and Anne R. Somers, *A Program for Research in Health Economics* (a background paper prepared for a conference held by the Brookings Institution), Washington, D.C., 1967, esp. pp. 37–39.

ing managing patients whom he cannot cure, preventing unnecessary exacerbation and complications, alleviating pain, and slowing down normal processes of decline. Listening to patients—taking their medical history—remains essential to the practice of good medicine, however scientific it may otherwise be.

Price Discrimination

The extraordinary growth of voluntary health insurance and the advent of Medicare have combined to accelerate the decline of price discrimination in medicine. Moreover, as Reder points out, its extent was probably overstated even in its prime.

Economists interested in studying medical care prices could contribute a great deal by inquiring into certain other aspects, such as the effects on price and quantity of fee schedules, of deductibles and coinsurance of varying amounts, of health insurance with varying degrees of coverage, of low physician incomes during prolonged periods of training, etc.

Reder's speculative discussion of the possible effects of the fee-for-service method of payment, such as fee splitting and inadequate allocation of time to postgraduate education, is imaginative. It strikes me as oversimple, in that there may be excessive surgery in the absence of fee splitting and conceivably also fee splitting without excessive surgery. As for postgraduate education, most specialists are paid fee-for-service and are held to keep up professionally. The reason lies in their strong hospital ties, not in taking time to attend formal courses.

REDER'S SCHEME

According to my understanding, Reder proposes that the following be done to measure change in the productivity of medical care.

1. Express the change as the ratio of change in physical output over the change in deflated inputs.

2. Define output preferably as services available to a subscriber to a comprehensive health insurance plan. Alternatively, in more realistic circumstances, express output in terms of the number of episodes of illness treated.

3. Adjust the above ratio by taking account of changes in quality. Five measures of quality are variously suggested: life years gained,

rates of undetected illness, appropriateness of the process of diagnosis and treatment, patient satisfaction, and work days lost. Marginal rates of substitution (trade offs) among the several measures of quality are to be obtained in a manner not specified.

CRITIQUE

In outline the scheme appeals to me as reasonable and workable. In detail I have a few doubts and several disagreements.

Unit of Output

It is important, I believe, that a position be taken on what constitutes output in the health services industry. Is it the services rendered by physicians and other personnel or is it the change effected in the health status of persons? In Reder's scheme the former is output in the first instance, while the latter qualifies as an adjustment in quality. I agree with this decision, because I can see no way of devising a proportionate relationship over time between health services and health status. Moreover, the ascertaining of such relationships is necessarily the object of special study, while the numerator of an index of productivity should be calculable on a fairly routine and reproducible basis.

What confuses me, for it is not explained, is Reder's apparent advocacy of two alternative measures of output that appear to be quite different.

Number of Subscribers. His preferred measure is the service available to a member of a comprehensive prepaid group practice plan. Although the quantity of services used by such a person may change over time, as technological possibilities change and consumer tastes and income change, the unit of output in Reder's scheme is the person eligible for care, standardized for certain characteristics, such as age and sex.

I can see no justification for the person as a measure of physical output (although he may be the appropriate unit for projecting medical care expenditures, for which changes in utilization are just as important as changes in price).

Episodes of Illness. The measure of output Reder considers second best but probably of greater applicability in the real world, in view

of the small numbers currently enrolled in prepaid group practice plans, is the number of episodes of illness treated. If the number of episodes treated per subscriber changes, the two measures of output are bound to differ.

Let me try to summarize the objections I can see to treating an episode of illness as the measure of output. Perhaps further discussion can serve to dispel them.

1. With the possible exception of some surgery, most illness cannot be divided over time into distinct episodes.

2. Indeed, even for a given diagnostic condition episodes of illness are variously accompanied by complications and multiple diseases, which may outlast the episode.

3. Certain health services are completely divorced from episodes of illness. A medical examination not prompted by symptoms is the most obvious example.

4. Certain health services are preventive in nature. Their purpose is to prevent the occurrence of illness. Vaccinations are an obvious example.

5. The management of chronic diseases cannot be divided into episodes.

If a major reason for taking the subscriber to a health plan as the unit of output is the fact that health services are increasingly purchased through an insurance premium, the obverse of this argument is that health services are not commonly paid for in terms of treatment for an episode of illness.

A Bundle of Services. Reder dismisses the present medical care component of the CPI, which prices specified bundles of services, as a possible measure of output and productivity, because no provision is made for taking account of changes in inputs. It seems to me, however, that it would be altogether in order to propose that this be done. If so, a specified basket of medical services emerges as another possible measure of output.

Although personally skeptical of the episode of illness approach, I do not believe we are ready to choose any single measure of output over the other two. More thinking is in order and certainly some hypothetical calculations and comparisons. It is conceivable that each

of the three measures offers certain advantages, depending on the problem at hand.

Adjustment for Quality

As for changes in the quality of care, I agree with Reder in making this a second step, an adjustment, rather than a direct measure of output.

The space given by Reder to quality is justified by its importance. He draws on a great deal of literature concerning the medical audit and quality assessment that economists too frequently consider outside their purview and repeatedly demonstrates its relevance. It strikes me, however, that Reder is perhaps less critical in his reference to this literature than he is likely to be in a similar review of work done by economists.

It will be helpful to recognize that the heading of quality usually encompasses several distinctive aspects: end results, process, and amenities.

I believe that it is always more meaningful to judge the quality of medical care by its effects on health status—whether measured in terms of gains in longevity, reduction of disability, or alleviation of pain—than by the conformity of the diagnostic and therapeutic process to prevailing criteria of good medical practice. At most, the latter criterion should be employed when information concerning end results is lacking. Hopefully, the lack of such information is temporary, for it serves as a signal that there is research to be done.

I have doubts concerning one of the suggested indicators of quality, the rate of undetected illness. In every instance it would be necessary to know what the cost of detection is in terms of false positives,[2] the natural history of the disease in the absence of intervention, and the availability of effective treatment for the detected condition. Even the desirability of early cancer detection programs (prior to symptoms) has been questioned recently on the ground that the natural history of preinvasive lesions is not known.[3]

[2] W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach,* Glencoe, Ill., 1956, pp. 328–329.

[3] E. G. Knox, "Cervical Cytology: A Scrutiny of the Evidence," in Gordon McLachlan (ed.), *Problems and Progress in Medical Care,* London, pp. 277–309.

It would be useful to consider as a measure of quality the absence of complications engendered by the medical treatment itself. Surely iatrogenic illness ranks first among all the conditions that should be avoided, as Florence Nightingale insisted many years ago.

In the *Milbank Quarterly,* a paper cited by Reder, Fuchs has shown that an economist can perform a highly acceptable review of the medical literature to distinguish between existing knowledge and claims of the efficacy of medical care. I would add that greater insistence on information concerning end results might serve both to accelerate research in this field and move epidemiologists away from their emphasis on experiments of classic design (study and control groups selected by random assignment). There is considerable room for rough and ready, relatively dirty methods for evaluating the end results of health services.[4]

Changes in the amenities appear to be easier to measure than end results. Improvements in the amenities of medical care are to be expected, along with other elements in the standard of living. It is time to recognize that biological need is not the only factor in the use of health services, and that greater conveniences or lesser discomforts, while not strictly necessary, do not represent abuses or wasteful use of health services.

Finally, with respect to quality, it would be very useful if Reder elaborated some possible approaches to the problem of rendering the several measures of quality commensurate. For example, to what extent can we draw on the work that has been done in estimating the benefits, tangible and intangible, of health service programs and perhaps go beyond it?

INPUTS

My comments about inputs are brief.

Reder has chosen to discuss the services rendered by a physician or under a physician's direction, rather than all personal health services, such as those rendered by dentists, nurses when acting autonomously,

[4] Peter H. Rossi, "Boobytraps and Pitfalls in the Evaluation of Social Action Programs," *Proceedings of the Social Statistics Section, American Statistical Association, 1966,* Washington, D.C., pp. 127–132; Sam Shapiro, "End Result Measurement of Quality of Medical Care," *The Milbank Memorial Fund Quarterly,* Vol. 45, No. 2 (April 1967, Part I), pp. 7–30.

or pharmacists. The reason for this limitation, it appears to me, is his undue preoccupation with health insurance and professional licensure as the two factors that define medical care. Medical care was provided long before the development of health insurance and would continue to be provided if the licensure of physicians were eliminated. Is it not sufficient to say that the health services industry comprises whatever responsible authorities say it is and serious students of the subject are willing to accept? Today, by virtue of pioneer work and persistent application, the Social Security Administration is the accepted authority, and the Internal Revenue Service need not agree with it, as it does not on such items as travel expenses. In studying specific problems the boundaries of the industry can be modified.

The fact that items of consumption other than medical care can and do affect health status is widely recognized. Medical care is one avenue for improving health status. Possibilities for making substitutions among means vary, depending on the particular disease entity. The importance of this point is the heavy burden that it puts on studies that evaluate end results. Not only is it necessary to ascertain the effects, but also to relate them to a source.

SOME APPLICATIONS

If Reder's general scheme is followed, certain problems that seemed intractable in the past appear to be on the road to a solution.

How to treat unnecessary surgery? Count the operation as output, but adjust it downward for poor quality. The adjustment can exceed the initial value of the output.

How to treat effective prevention? Clearly output in later years is reduced. Output in the early years will be adjusted upward for quality, most likely retroactively, in light of the accumulating evidence.

How to treat costless improvements in medical care, such as early ambulation? Output is reduced, but the gain in quality is substantial. The latter is greater if early discharge also signifies an early return to normal activity. Input is zero or small (the latter is due to heightened activity over a shorter duration). If costs were incurred in research, they belong there and not in inputs for services. There is no

reason to reward the producers of services for such an improvement in quality, which is costless to them and was paid for, if at all, in the research sector.

SUMMARY

To summarize: we have here a thoughtful, wide-ranging, and stimulating discussion of a major subject area in health economics. I believe that Reder's principal decisions are sound.

My major reservation in the economist's sphere relates to the choice of a measure of output. My disagreements reflect largely differences in interpreting the current status of the literature on quality. I invite Reder's further efforts in devising approaches to render the several dimensions of quality commensurate. Finally, possible applications of Reder's approach are indicated.

MARTIN S. FELDSTEIN, Harvard University

As you all know, Professor Reder's paper offers much more than his title suggests. Reder begins by discussing a number of alternative approaches to the measurement of productivity change in the medical care industry, but then goes on to speculate about some of the ways in which the organization of the industry influences the productivity of medical manpower.

Because my time is short, I will restrict most of my comments to the problems of productivity measurement. But first I would like to call attention to some of the aspects of Reder's more general discussion that I found particularly interesting.

Much of Reder's attention focuses on the influence that the method of paying physicians has on their behavior. A novel feature of his analysis is the treatment of the physician payment method as a form of risk-sharing arrangement. Other aspects of his discussion are also unusual. Although Reder supports the widespread criticism that fee-for-service remuneration provides little or no incentive for "high-quality" medicine while encouraging unnecessary surgery and other improper treatment, his most interesting criticism is that doctors are

generally not responsive enough to the potential market demands of patients. In particular, Reder speculates that patients would be willing to pay more for the treatment of miscellaneous psychosomatic illnesses and for greater convenience in their care. In spite of the fee-for-service payment method, patients' potential demand is ineffective because doctors are not income maximizers.

Reder also develops the idea that the market mechanism is currently a very imperfect allocator of medical services because potential patients are unwilling to spend the time required to obtain adequate information about the quality of available services. Reder's emphasis on patients' reluctance to spend time, rather than on their inability to understand technical information, has an important policy implication. It suggests that the government or private organizations might be able to provide the public with information that would improve the efficiency of the medical market. I would like to hear Professor Reder expand this point.

Although there are other interesting and potentially fruitful ideas in Reder's discussion of the factors influencing physician productivity, I want to return now to a more detailed analysis of his treatment of productivity measurement.

Reder correctly emphasizes that the measurement of price and productivity change should be in terms of the outputs of the medical care industry and not its inputs. He therefore rejects the method currently used in calculating the medical component of the consumer price index. In the search for a substitute, he discusses several different approaches to measuring the "output" and the rate of productivity change of the medical care industry. I will consider three of these.

The first and most striking of Reder's suggested approaches is his "cost of insurance index." Reder starts with the harmless convention of defining the medical care industry to be coterminous with the services provided under a broad medical insurance or prepayment plan. He then jumps to the much stronger assertion that: "If medical care is that which can be purchased by means of medical care insurance, then its 'price' varies proportionately with the price of such insurance." With these definitions of output and price, Reder suggests that changes in resource productivity in the production of medical care might be measured by the relative changes in an index of medical

insurance premiums and an index of input prices. If the cost of insurance remains constant while input prices rise, productivity is said to increase; if insurance premiums rise with given input prices, productivity is said to fall.

I find the "cost of insurance" approach to price and productivity measurement unsatisfactory. It fails to distinguish between *unit* price and *total* (expected) expenditure. An individual's total expenditure on medical care reflects the number of services purchased, the average "quality" of each service, and the prices of services of different qualities. If all unit prices remain unchanged between two successive years, but consumers purchase more services or services of higher quality, the "cost of insurance" price index will rise and the implied productivity change will be negative.

In practice the "cost of insurance" price index is almost certain to be biased upward. If the calculations are not restricted to "fully comprehensive" insurance programs, average premiums will rise through time in reflection of the trend toward more comprehensive coverage. But even if attention were restricted to fully comprehensive insurance, I suspect that the number and average quality of services consumed would increase with time in response to technical innovation, increased availability, and higher aspirations on the part of doctors and patients. Some of my own research on the operation of the British National Health Service [1] supports this view. Although everyone in Britain is legally covered by a free and comprehensive medical care system, substantial differences in the utilization of services are associated with the availability of facilities, with the social class of patients, and with the treatment preferences of general practitioners. The effect of such an upward trend in utilization would be an upward bias in the price index and a downward bias in measured productivity.

Reder carefully qualifies his discussion by stating that the "cost of insurance" method of productivity measurement is only valid if the "quality" of care is unchanged. Because he defines "quality" of care very broadly—to mean both the *number* of services consumed and the *average quality* per service—he is guilty of no formal error. But it

[1] M. S. Feldstein, *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service,* Amsterdam, 1967.

seems to me that this merely hides the basic difficulty of *aggregating* individual health sector outputs under the more familiar heading of quality measurement.

Reder's second approach to output measurement is to use a disaggregated description of the physical health benefits produced in each year or, if cross-section comparisons are to be made, in each patient group. The measures of physical health benefits would include not only mortality and morbidity rates but also assessments of patient satisfaction, the number of undetected illnesses, etc. Although such an approach appeals to me as one of substantial promise, it faces two serious difficulties.

The first of these is *estimating* the extent to which differences in medical care are responsible for differences in health variables, either over time or between groups. It might at first appear that such estimates could be provided by a set of multiple regression equations, each relating one aspect of the health level (e.g., a particular mortality rate) to variables which measure the inputs of medical services and the population characteristics relevant to the dependent variable. Several researchers have in fact estimated equations of this type. In practice, high correlation among the explanatory variables generally makes it impossible to obtain estimates with any precision; the more scrupulous the researcher is in including potentially relevant explanatory variables, the more serious the multicollinearity problem becomes. But the least squares regression approach is subject to more serious criticism on the theoretical level. First, it implies that the community's health at the present time depends only on the *current* levels of medical care inputs and not on the previous use of health services. A more sophisticated analytic framework, allowing for different delayed responses to medical inputs and environmental factors, would therefore be necessary. Second, the single equation regression ignores the fact that the levels of medical care inputs are determined partly in response to the health of the community. Since areas with a naturally healthier population will tend to use fewer medical services, the regression approach will underestimate the benefits of increasing medical care inputs. Third, to calculate the effect of medical care on any one aspect of community health, it is necessary to hold constant statistically all of the other outputs of medical care. Failure to do so would cause a further downward bias in the estimated effect

of medical inputs. As you can see, estimating the extent to which differences in medical care inputs are responsible for differences in health variables poses a number of difficult econometric problems. But the usefulness of such estimates for general health sector planning as well as for productivity measurement, makes the solution of these problems an important research task.

The second difficulty inherent in Reder's "physical benefit" approach is the problem of aggregating the individual physical benefits to obtain a composite scalar output variable. Reder suggests that this be done by defining a set of weights reflecting the marginal rates of substitution between outputs. This is essentially what is done for consumer goods in general by using relative market prices. But there the output is carefully defined as an index of *market value* at constant prices. But Reder's index seems to me to be intended as a *total* welfare or *satisfaction* measure of output rather than a market value output measure. The use of *constant* marginal rate of substitution weights implies that the underlying utility or welfare function is linear. But the appropriate marginal rate of substitution between, say, morbidity from disease A and morbidity from disease B might well depend upon the current levels of both morbidity rates and upon other dimensions of health as well. If this is so—if the appropriate utility function is neither linear nor even separable—it would be very difficult to obtain the information about preferences that is needed to evaluate the *total* output of the health care industry. One possible solution to this problem is to abandon the attempt to evaluate *total* output and to limit analysis to the *changes* in output. For small changes from a previous value and for small differences between groups, constant marginal-rate-of-substitution weights would be suitable as a linear approximation to the utility function at its current levels of output. Although the lack of a value for total output would preclude any estimate of the *proportional* output change, it would be possible to compare output changes between successive pairs of years, etc.

The third and final approach to productivity measurement that I shall consider is Reder's adaptation of Mrs. Scitovsky's study of the costs of specific illnesses.[2] If attention is restricted to an illness for

[2] A. A. Scitovsky, "Changes in the Costs of Treatment of Selected Illnesses, 1951–65." *American Economic Review,* December 1967, pp. 1182–1195.

which the effectiveness of treatment has not varied over time, productivity change may be defined as the reciprocal of an index of the value of inputs per case. A productivity index for all the illnesses in which treatment effectiveness remained unchanged could then be defined by a suitable weighting of these individual productivity measures. The obvious serious shortcoming of this method is that attention must be restricted to illnesses in which treatment effectiveness has not changed.

Let us look at the reasons why the use of this method is likely to be misleading—both as a measure of what Reder called "productivity change for . . . quality constant output" and as a basis for judging how "technologically progressive" the medical care industry is. In general, changes in the productivity of treating an illness may be thought of as reflecting two things: changes in the efficiency with which resources are used with a given technology, and innovations which change the technology itself. Such innovations may be divided into product innovations (i.e., those which alter the effectiveness of treatment) and process innovations (i.e., those which leave the effectiveness of treatment unchanged while reducing cost). In these terms, the Reder-Scitovsky method (if I may call it that without implying that it is advocated by either of them) may be described as trying to estimate the effects of process innovation and increased efficiency of resource use by restricting attention to those illnesses in which there has been no product innovation. The method will only be successful if the rate of product innovation is independent (across illnesses) of the rate of process innovation and of the change in the efficiency of resource use. If, for example, illnesses with positive rates of product innovation also have higher rates of process innovation, the Reder-Scitovsky index will be biased downwards as a measure of the entire industry's "productivity change . . . for quality constant output." A bias in the opposite direction is more likely to occur because the absence of product innovation probably facilitates increasing the efficiency with which the unchanged product is produced. The sign of the net bias is nevertheless indeterminate. However, by pursuing this problem further, we might be able to learn more about both productivity measurement and the nature of technical progress in the health care field.

Although the Reder-Scitovsky method may provide a useful and improvable measure of productivity change for "quality constant output," it is wholly inadequate as a basis for judging how "technologically progressive" the medical care industry is. Product innovation is probably the most important aspect of its technological progressiveness. The ability to achieve high rates of product innovation must therefore be heavily weighted in any policy assessment of the organization of health care services.

That concludes my comments on Reder's suggested methods of productivity measurement. My own view is that medical care output may legitimately and usefully be defined in any of four ways: by an index of the number of services provided, of the number of cases treated, of the number of successful treatments, or of various measures of the community's health. Each definition is progressively more difficult to implement than the preceding one but comes closer to what we want for welfare-oriented comparisons of output.[8]

There is also a further dimension of output that should not be overlooked: uncertainty. Anything that increases the probability that effective care will be available when requested should be counted as an improvement in the quality of output. As per capita income increases, there will be a greater willingness to pay for reduced uncertainty. Such reductions will be achieved, at least in part, by increases in excess capacity. As a result, costs will rise without a concomitant increase in tangible output. But it would be as wrong to consider this to be a fall in productivity as it would be to measure the productivity of a fire department by the number of fires extinguished per fireman. A failure to allow for the contribution of excess capacity to the reduction of uncertainty will bias downward the measured productivity change.

Although my remarks have generally been critical, I think that Reder's contribution is a substantial one. It provides a bold depar-

[8] The dividing line between inputs and outputs is unclear. For me, the distinguishing characteristic is the possibility of substitution. Although Reder classifies a hospital bed-day as an input, I would treat it as a form of output because it can be produced with different combinations of inputs. I would not deny, of course, that a bed-day is also an input used in producing the output, "a treated case." But a treated case is also an input in producing an improvement in the community's health level.

ture from the traditional national income accounting methods of measuring price and productivity change. Medical care is not like most other goods; Reder has shown us that we must be prepared to use novel methods to deal with this new problem.

COMMENTS

SOLOMON FABRICANT, New York University

Melvin Reder has put before us some interesting and instructive points which we all need to keep in mind when we worry about measurements of the output, price and productivity of the medical-care industries.

One suggestion that crops up bothers me enough, however, to force an immediate reaction. This is the suggestion that we bypass a lot of problems and measure the productivity of the medical-care industries by the reciprocal of the real cost incurred in maintaining a given state of the health of the population.

Were "all other things" the same or of negligible importance, this would make good sense. But neither is true. Health is a function not only of the output of the medical-care industries but also of many other industries and activities—education in hygiene, refrigeration, meat inspection, sports, water and sewage systems, etc. Can we really argue that medical care constitutes the major factor in maintaining health, and that the other factors are of negligible importance? Or that they have changed exactly in proportion to the former—measured in terms of input? There are many reasons to believe the contrary. The other factors are not unimportant, nor have they grown in exact proportion to medical-care output as economic development has proceeded.

For some purposes, perhaps, we are concerned not with the sources of good health but only with the level and distribution of health. As economists, however, we do require knowledge of these sources and their return per dollar of expenditure. Shall municipal or state or federal authorities—or the individual family—spend more or less on doctors, mosquito eradication or screening, better food, house hu-

midifiers, or the multitude of other things that affect health? To answer these questions, we must do the best we can to measure and compare the marginal revenue yielded by the last dollar spent on each type of expenditure. This requires, among other things, measuring the output of the health service industries in some more specific and narrower way than by the general level of health. We are back to the original question.

Since we cannot skirt the difficulties, we must meet them head-on. I expect that when we do, measuring not only the effect on health but also the other benefits (or costs) of medical care, we will profit from many of the points made by Reder. I expect also that we will find the trend in the output and productivity of the medical-care industries growing, and the price falling, much more than current measurements indicate. Finally, I suspect that we will discover enough variation in the degree to which there is need to correct these measurements in different periods and places, as to make untenable the assumption that the bias can be neglected.

WERNER Z. HIRSCH, University of California at Los Angeles

In developing a production function for the medical-care industry it might be useful to separate the independent variables into two groups. On the one hand, there are the conventional input factors, such as different types of labor, capital and material. On the other hand, there are service conditions affecting input requirements which, in a sense, are input factors with negative effects. Perhaps a more useful distinction between conventional input factors and servicing conditions affecting input requirements is that by and large the former can be influenced, and perhaps even controlled, by the decision maker, while this is not true about the latter.

In considering service conditions affecting input requirements, we can point to a variety of physical, human, financial, legal and political factors which can make it easier, or more difficult, for medical-care services to be rendered. Specific variables to be considered are age of population, nutrition, housing conditions, climate, etc.

REPLY BY REDER

Had I been sufficiently prescient to anticipate the observations of my critics, this paper would have been somewhat different both in exposition and in content. I have made some minor changes from the preliminary version both in response to the remarks of Feldstein and Klarman, and to the suggestions of a number of others. However, I have tried to avoid such alterations as would have the effect of removing the target of a discussant's shaft.

CONCEPT OF OUTPUT

Prepaid Care

While definitions are arbitrary, their analytical consequences are not. To define a unit of medical care as that which would be offered an individual under a nationwide comprehensive prepayment plan entails that output must vary proportionately with the number of individuals covered, and that there is no way in which output per standardized individual can change.[1] All of what has hitherto been considered a variation in output quantity per individual—usually an implicit inference from a variation in one or more input quantities—must, on this definition, be considered as a variation of output quality.

One of the reasons why Martin Feldstein objects to this proposed output concept is precisely that it has this consequence. I readily concede the advantage of an output measure that makes it possible to distinguish between changes in the number of service units consumed (per individual) per time period and changes in their quality. However, to define the service units rendered a given standardized individual in a way that permits this distinction causes other difficulties that are at least as serious. For example: (1) to measure output quan-

[1] "Standardized individual" refers to the number of physical persons weighted by a measure of their expected use of medical services at given prices and income. Expected use of medical services is conditional upon the age, sex and various other characteristics of an individual. In principle, standardization is supposed to eliminate the effect of changes in the composition of a population (with respect to age, sex, etc.) upon its consumption of medical service.

tity by the quantities of any set of inputs, or any function thereof, is to preclude the possibility of observing any variation in the efficiency with which inputs are used to generate outputs; (2) to use episodes of illness treated as the unit of output is open to the objections that Klarman makes so forcefully, as well as to still others.

Defining output of medical care as the number of standardized individuals cared for under comprehensive prepaid coverage, presents many practical problems of implementation, as Feldstein indicates. Care must be truly comprehensive; patient inconvenience or unavailability of service must be considered either as an unpaid input or a quality variation. Lagged effects of past treatment must also be allowed for; a person with a twenty-year history of "good care," will presumably need less treatment than one of the same age and sex who has systematically "under-doctored." In principle, the appropriate adjustment should be made in the standardization process; in practice, rough and ready ad hoc adjustments will have to be made. In inferring changes in quality of care from changes in health conditions of individuals, it is essential also to make adjustments for changes in environmental factors (e.g., change in air pollution) that affect health.[2]

Obviously, it is much easier to say "make allowance" for one extraneous influence or another than to do it. In practice, most of the suggested adjustments and allowances will have to be performed in a quite crude manner. However, the difficulties arise mainly from the attempt simultaneously to measure output directly (not accepting one or more inputs as a surrogate measure) and to estimate the share of changes in measured productivity attributable to changes in the quality of product. Similar difficulties will arise whatever the measure of output chosen, provided that measurements of output change are to be net of adjustments for quality change. It is not clear that the particular output measure discussed involves greater difficulties than alternatives suggested thus far.

[2] I interpret the comments of both Solomon Fabricant and Werner Hirsch to be essentially to this same point, and I agree with them. Health depends on many things beside medical care and I share, what I take to be, Fabricant's view that improvements in these other factors may have been more important in promoting health than those in medical care. But to analyze the effect of productivity change in the provision of medical care upon health, it is necessary to adjust for the effect of these other variables on health. To make such adjustments is not to minimize the importance of the variables from whose influence we abstract.

Fee for Service

Though critical, Feldstein views the output measure relevant to fee for service treatment more favorably than does Klarman. Klarman properly objects to this measure as ignoring the role of preventive care. Possibly the objection could be overcome by including in output a specified standard amount of diagnostic care, the amount varying with the age and sex of the individual in question.[3] Up to the specified amount, diagnostic care would be counted as output; beyond it, such care would not be so considered.[4] For specific individuals, such an assumption might cause substantial error, but such errors should cancel out for a large group. Nevertheless, I readily concede the appropriateness of Klarman's objection.

Feldstein offers two objections: (1) there are serious difficulties in estimating the effect of current medical care on health. Certainly, I agree that estimation of changes in the efficiency of medical care should take into account lagged response to past treatment. Feldstein is also correct in warning against the dangers of single equation bias. As I interpret him on this point, what he is arguing is that an apparent increase in inputs required for treating a particular ailment for a typical member of a given age-sex class may appear to reflect a decline in productivity, but may be due in fact to an increase in the normal severity of the ailment because of environmental changes. In principle, such confusions can be avoided by including one or more variables reflecting environmental conditions in the regression. In practice, specifying the appropriate variables is hard, and one can never be very confident that he has caught all of the effects of environmental change or of changes in "taste for medical care." However, the only pertinent advice is "be careful."

(2) Feldstein objects to my suggested aggregation procedure for physical benefits because he supposes I am trying to define a welfare index of some sort. My purpose is quite different: in my opinion, welfare judgments as to the marginal rate of substitution between,

[3] The specified amount would be determined by an expert panel and would represent a standard for the typical member of an appropriate age-sex-environment class.

[4] This suggestion involves fudging productivity measurement since it measures output of care by input of diagnostic services.

say, a reduction in the incidence of one ailment vis-à-vis another are simply statements about the properties of the judge's individual preference function over the set of possible states of affairs.

It is not necessary that we specify whether or how these individual preference functions are aggregated to arrive at the requisite trade-off relations; for the purpose of this paper, the specifications of these trade-offs are assumed. In practice, the specification might be made as an aspect of a program specifying "national goals" for research and delivery of health care; i.e., they might reflect the preferences of a national health commission. There are a number of other ways by which the requisite characteristics might be posited or even inferred from collective behavior but this is too large a topic to go into here.

In discussing the aforementioned trade-offs (marginal rates of substitution) I did not mean to suggest that they were constants; only that they were specified at each state of "national health." I suspect that it was my inadvertent use of the term "weights" (for the marginal rates of substitution) that misled Feldstein on this point. Variable "weights" create problems for constructing output and productivity indexes precisely analogous to those created by varying prices, i.e., which set of prices should be used as weights. These problems can be solved as well or as poorly as their more familiar analogues.

I also agree with what I take to be the thrust of Feldstein's comment upon uncertainty. However, I believe the main force of his remarks (on this point) bears upon the costs of queueing and of delivery delay rather than upon uncertainty per se.[5] Feldstein is correct in arguing that it is desirable to treat variations in reserve capacity for furnishing medical service as a variation in output quality. However, it should be noted that his remarks apply to a very large class of industries, both in the service and nonservice sectors, and not only to medical care. Conventional output measures fail to allow for the productive role of reserve capacity because they take no account of the economic cost of delivery delay, i.e., the costs of queueing are neglected. Thus the effect of a reduction in expected delivery

[5] One very important reason for building reserve capacity is to hedge against uncertainty. However indivisibilities may also lead to (voluntary) holdings of reserve capacity in the absence of uncertainty. Such reserve capacity may well serve to reduce average delivery delays even though they are perfectly foreseen.

delay because of an increase in reserve productive capacity may suggest reduced efficiency because it involves increased pecuniary outlay (which is recorded) while its yield (delivery convenience) goes unmeasured.

Feldstein is again correct in warning us of the dangers of estimation bias inherent in using information which refers only to ailments where treatment effectiveness has been invariant as indicative of productivity change for all ailments. I intended footnote 19 as a warning to the same effect but apparently its message did not get through.

PRICE DISCRIMINATION

Klarman's statement that price discrimination is of declining importance may be misleading. Price discrimination between richer and poorer patients probably is declining. But price discrimination between desirable and undesirable patients, whether for personal or diagnostic reasons, may be on the increase; i.e., use of standard fee schedules may disguise wide variations in time and attention offered for a given fee. This is clearly a matter that needs investigation.

CRITERIA FOR JUDGING QUALITY

I fear that Klarman has detected in me a residue of layman respect for medical expertise. He may be right in accusing me of being insufficiently critical of the literature on medical audit. In defense, I plead technical ignorance. Klarman's long experience in the health field gives him a justified independence of judgment in these matters vouchsafed to few economists. It is very desirable that he write a critique of medical audit procedures for the benefit of us social scientists who are neophytes in the health care field.

However, Klarman's appeal for "rough and ready, relatively dirty methods" for evaluating end results of health services disturbs me. This may be sound procedural advice to epidemiologists with great experience in the field, but it would be disastrous for economists to apply such methods on their own. In emphasizing "experiments of classic design," epidemiologists may be exhibiting too much caution, but they may also be indicating fear of incorrect specification of their models. The impatience of men of affairs, combined with our own desire to provide immediately useful results, has caused economics

many embarrassing episodes. We should be wary of foisting on other disciplines the methodological errors we are outgrowing in our own.

Klarman's reservations concerning the rate of undetected illness as an (inverse) quality indicator seem reasonable. However, if all I knew about two physicians was their "undetected illness rates," I would choose the one whose rate was lower.