

THE AMERICAN ECONOMIC REVIEW

VOLUME LIII

DECEMBER 1963

NUMBER 5

UNCERTAINTY AND THE WELFARE ECONOMICS OF MEDICAL CARE

By KENNETH J. ARROW*

I. Introduction: Scope and Method

This paper is an exploratory and tentative study of the specific differentia of medical care as the object of normative economics. It is contended here, on the basis of comparison of obvious characteristics of the medical-care industry with the norms of welfare economics, that the special economic problems of medical care can be explained as adaptations to the existence of uncertainty in the incidence of disease and in the efficacy of treatment.

It should be noted that the subject is the *medical-care industry*, not *health*. The causal factors in health are many, and the provision of medical care is only one. Particularly at low levels of income, other commodities such as nutrition, shelter, clothing, and sanitation may be much more significant. It is the complex of services that center about the physician, private and group practice, hospitals, and public health, which I propose to discuss.

The focus of discussion will be on the way the operation of the medical-care industry and the efficacy with which it satisfies the needs of society differ from a norm, if at all. The "norm" that the economist usually uses for the purposes of such comparisons is the operation of a competitive model, that is, the flows of services that would be offered and purchased and the prices that would be paid for them if each individual in the market offered or purchased services at the going prices as if his decisions had no influence over them, and the going prices were such that the amounts of services which were available equalled the total amounts which other individuals were willing to purchase, with no imposed restrictions on supply or demand.

The interest in the competitive model stems partly from its presumed descriptive power and partly from its implications for economic efficiency. ...

If, ... , the actual market differs significantly from the competitive model, or if the assumptions of the two optimality theorems are not fulfilled, the separation of allocative and distributional procedures becomes, in most cases, impossible.

The first step then in the analysis of the medical-care market is the comparison between the actual market and the competitive model. ...

* The author is professor of economics at Stanford University. He wishes to express his thanks for useful comments to F. Bator, R. Dorfman, V. Fuchs, Dr. S. Gilson, R. Kessel, S. Mushkin, and C. R. Rorem. This paper was prepared under the sponsorship of the Ford Foundation as part of a series of papers on the economics of health, education, and welfare.

In this paper, the institutional organization and the observable mores of the medical profession are included among the data to be used in assessing the competitiveness of the medical-care market. I shall also examine the presence or absence of the preconditions for the equivalence of competitive equilibria and optimal states. The major competitive preconditions, in the sense used here, are three: the *existence* of competitive equilibrium, the *marketability* of all goods and services relevant to costs and utilities, and *nonincreasing returns*. The first two, as we have seen, insure that competitive equilibrium is necessarily optimal; the third insures that every optimal state is the competitive equilibrium corresponding to some distribution of income.⁴ The first and third conditions are interrelated; indeed, nonincreasing returns plus some additional conditions not restrictive in a modern economy imply the existence of a competitive equilibrium, i.e., imply that there will be some set of prices which will clear all markets. ...

That risk and uncertainty are, in fact, significant elements in medical care hardly needs argument. I will hold that virtually all the special features of this industry, in fact, stem from the prevalence of uncertainty.

The nonexistence of markets for the bearing of some risks in the first instance reduces welfare for those who wish to transfer those risks to others for a certain price, as well as for those who would find it profitable to take on the risk at such prices. But it also reduces the desire to render or consume services which have risky consequences; in technical language, these commodities are complementary to risk-bearing. Conversely, the production and consumption of commodities and services with little risk attached act as substitutes for risk-bearing and are encouraged by market failure there with respect to risk-bearing. Thus the observed commodity pattern will be affected by the nonexistence of other markets.

The failure of one or more of the competitive preconditions has as its most immediate and obvious consequence a reduction in welfare below that obtainable from existing resources and technology, in the sense of a failure to reach an optimal state in the sense of Pareto. But more can be said. I propose here the view that, when the market fails to achieve an optimal state, society will, to some extent at least, recognize the gap, and nonmarket social institutions will arise attempting to bridge it. Certainly this process is not necessarily conscious; nor is it uniformly successful in approaching more closely to optimality when the entire range of consequences is considered. It has always been a favorite activity of economists to point out that actions which on their face achieve a desirable goal may have less obvious consequences, particularly over time, which more than offset the original gains.

But it is contended here that the special structural characteristics of the medical-care market are largely attempts to overcome the lack of optimality due to the nonmarketability of the bearing of suitable risks and the imperfect marketability of information. These compensatory institutional changes, with some reinforcement from usual profit motives, largely explain the observed noncompetitive behavior of the medical-care market, behavior which, in itself, interferes with optimality. The social adjustment towards optimality thus puts obstacles in its own path. ...

II. A Survey of the Special Characteristics of the Medical-Care Market

This section will list selectively some characteristics of medical care which distinguish it from the usual commodity of economics textbooks. The list is not exhaustive, and it is not claimed that the characteristics listed are individually unique to this market. But, taken together, they do establish a special place for medical care in economic analysis.

A. The Nature of Demand

The most obvious distinguishing characteristics of an individual's demand for medical services is that it is not steady in origin as, for example, for food or clothing, but irregular and unpredictable. Medical services, apart from preventive services, afford satisfaction only in the event of illness, a departure from the normal state of affairs. It is hard, indeed, to think of another commodity of significance in the average budget of which this is true. A portion of legal services, devoted to defense in criminal trials or to lawsuits, might fall in this category but the incidence is surely very much lower (and, of course, there are, in fact, strong institutional similarities between the legal and medical-care markets.)

In addition, the demand for medical services is associated, with a considerable probability, with an assault on personal integrity. There is some risk of death and a more considerable risk of impairment of full functioning. In particular, there is a major potential for loss or reduction of earning ability. The risks are not by themselves unique; food is also a necessity, but avoidance of deprivation of food can be guaranteed with sufficient income, where the same cannot be said of avoidance of illness. Illness is, thus, not only risky but a costly risk in itself, apart from the cost of medical care.

B. Expected Behavior of the Physician

It is clear from everyday observation that the behavior expected of sellers of medical care is different from that of business men in general. These expectations are relevant because medical care belongs to the category of commodities for which the product and the activity of production are identical. In all such cases, the customer cannot test the product before consuming it, and there is an element of trust in the relation.¹⁴ But the ethically understood restrictions on the activities of a physician are much more severe than on those of, say, a barber. His behavior is supposed to be governed by a concern for the customer's welfare which would not be expected of a salesman. In Talcott Parsons's terms, there is a "collectivity-orientation," which distinguishes medicine and other professions from business, where self-interest on the part of participants is the accepted norm.

... C. Product Uncertainty

Uncertainty as to the quality of the product is perhaps more intense here than in any other important commodity. Recovery from disease is as unpredictable as is its incidence. In most commodities, the possibility of learning from one's own experience or that of others is strong because there is an adequate number of trials. In the case of severe illness, that is, in general, not true; the uncertainty due to inexperience is added to the intrinsic difficulty of prediction. Further, the amount of uncertainty, measured in terms of utility variability, is certainly much greater for medical care in severe cases than for, say, houses or automobiles, even though these are also expenditures sufficiently infrequent so that there may be considerable residual uncertainty.

Further, there is a special quality to the uncertainty; it is very different on the two sides of the transaction. Because medical knowledge is so complicated, the information possessed by the physician as to the consequences and possibilities of treatment is necessarily very much greater than that of the patient, or at least so it is believed by both parties. Further, both parties are aware of this informational inequality, and their relation is colored by this knowledge.

D. *Supply Conditions*

In competitive theory, the supply of a commodity is governed by the net return from its production compared with the return derivable from the use of the same resources elsewhere. There are several significant departures from this theory in the case of medical care.

Most obviously, entry to the profession is restricted by licensing. Licensing, of course, restricts supply and therefore increases the cost of medical care. It is defended as guaranteeing a minimum of quality. Restriction of entry by licensing occurs in most professions, including barbering and undertaking. ...

E. *Pricing Practices*

The unusual pricing practices and attitudes of the medical profession are well known: extensive price discrimination by income (with an extreme of zero prices for sufficiently indigent patients) and, formerly, a strong insistence on fee for services as against such alternatives as prepayment. ...

III. *Comparisons with the Competitive Model under Certainty*

A. *Nonmarketable Commodities*

As already noted, the diffusion of communicable diseases provides an obvious example of nonmarket interactions. But from a theoretical viewpoint, the issues are well understood, and there is little point in expanding on this theme. (This should not be interpreted as minimizing the contribution of public health to welfare; there is every reason to suppose that it is considerably more important than all other aspects of medical care.)

Beyond this special area there is a more general interdependence, the concern of individuals for the health of others. The economic manifestations of this taste are to be found in individual donations to hospitals and to medical education, as well as in the widely accepted responsibilities of government in this area. The taste for improving the health of others appears to be stronger than for improving other aspects of their welfare.

In interdependencies generated by concern for the welfare of others there is always a theoretical case for collective action if each participant derives satisfaction from the contributions of all.

B. Increasing Returns

Problems associated with increasing returns play some role in allocation of resources in the medical field, particularly in areas of low density or low income. Hospitals show increasing returns up to a point; specialists and some medical equipment constitute significant indivisibilities. In many parts of the world the individual physician may be a large unit relative to demand. In such cases it can be socially desirable to subsidize the appropriate medical-care unit. The appropriate mode of analysis is much the same as for water-resource projects. Increasing returns are hardly apt to be a significant problem in general practice in large cities in the United States, and improved transportation to some extent reduces their importance elsewhere.

C. Entry

The most striking departure from competitive behavior is restriction on entry to the field, as discussed in II.D above. Friedman and Kuznets, in a detailed examination of the pre-World War II data, have argued that the higher income of physicians could be attributed to this restriction. ...

D. Pricing

The pricing practices of the medical industry (see II.E above) depart sharply from the competitive norm. As Kessel [17] has pointed out with great vigor, not only is price discrimination incompatible with the competitive model, but its preservation in the face of the large number of physicians is equivalent to a collective monopoly. In the past, the opposition to prepayment plans has taken distinctly coercive forms, certainly transcending market pressures, to say the least. ...

IV. Comparison with the Ideal Competitive Model under Uncertainty

A. Introduction

In this section we will compare the operations of the actual medical-care market with those of an ideal system in which not only the usual commodities and services but also insurance policies against all conceivable risks are available. Departures consist for the most part of insurance policies that might conceivably be written, but are in fact not. Whether these potential commodities are nonmarketable, or, merely because of some imperfection in the market, are not actually marketed, is a somewhat fine point.

To recall what has already been said in Section I, there are two kinds of risks involved in medical care: the risk of becoming ill, and the risk of total or incomplete or delayed recovery. The loss due to illness is only partially the cost of medical care. It also consists of discomfort and loss of productive time during illness, and, in more serious cases, death or prolonged deprivation of normal function. From the point of view of the welfare economics of uncertainty, both losses are risks against which individuals would like to insure. The nonexistence of suitable insurance policies for either risk implies a loss of welfare.

... C. *Problems of Insurance*

1. *The moral hazard.* The welfare case for insurance policies of all sorts is overwhelming. It follows that the government should undertake insurance in those cases where this market, for whatever reason, has failed to emerge. Nevertheless, there are a number of significant practical limitations on the use of insurance. It is important to understand them, though I do not believe that they alter the case for the creation of a much wider class of insurance policies than now exists.

One of the limits which has been much stressed in insurance literature is the effect of insurance on incentives. What is desired in the case of insurance is that the event against which insurance is taken be out of the control of the individual. Unfortunately, in real life this separation can never be made perfectly. The outbreak of fire in one's house or business may be largely uncontrollable by the individual, but the probability of fire is somewhat influenced by carelessness, and of course arson is a possibility, if an extreme one. Similarly, in medical policies the cost of medical care is not completely determined by the illness suffered by the individual but depends on the choice of a doctor and his willingness to use medical services. It is frequently observed that widespread medical insurance increases the demand for medical care. Coinsurance provisions have been introduced into many major medical policies to meet this contingency as well as the risk aversion of the insurance companies.

To some extent the professional relationship between physician and patient limits the normal hazard in various forms of medical insurance. By certifying to the necessity of given treatment or the lack thereof, the physician acts as a controlling agent on behalf of the insurance companies. Needless to say, it is a far from perfect check; the physicians themselves are not under any control and it may be convenient for them or pleasing to their patients to prescribe more expensive medication, private nurses, more frequent treatments, and other marginal variations of care. It is probably true that hospitalization and surgery are more under the casual inspection of others than is general practice and therefore less subject to moral hazard; this may be one reason why insurance policies in those fields have been more widespread.

2. *Alternative methods of insurance payment.* It is interesting that no less than three different methods of coverage of the costs of medical care have arisen: prepayment, indemnities according to a fixed schedule, and insurance against costs, whatever they may be. In prepayment plans, insurance in effect is paid in kind—that is, directly in medical services. The other two forms both involve cash payments to the beneficiary, but in the one case the amounts to be paid involving a medical contingency are fixed in advance, while in the other the insurance carrier pays all the costs, whatever they may be, subject, of course, to provisions like deductibles and coinsurance. ...

The need for third-party control is reinforced by another aspect of the moral hazard. Insurance removes the incentive on the part of individuals, patients, and physicians to shop around for better prices for hospitalization and surgical care. The market forces, therefore, tend to be replaced by direct institutional control.

6. *Pooling of unequal risks.* Hypothetically, insurance requires for its full social benefit a maximum possible discrimination of risks. Those in groups of higher incidences of illness should pay higher premiums. In fact, however, there is a tendency to equalize, rather than to differentiate, premiums, especially in the Blue Cross and similar widespread schemes. This constitutes, in effect, a redistribution of income from those with a low propensity to illness to those with a high propensity. The equalization, of course, could not in fact be carried through if the market were genuinely competitive. Under those circumstances, insurance plans could arise which charged lower premiums to preferred risks and draw them off, leaving the plan which does not discriminate among risks with only an adverse selection of them.

As we have already seen in the case of income redistribution, some of this may be thought of as insurance with a longer time perspective. If a plan guarantees to everybody a premium that corresponds to total experience but not to experience as it might be segregated by smaller subgroups, everybody is, in effect, insured against a change in his basic state of health which would lead to a reclassification. This corresponds precisely to the use of a level premium in life insurance instead of a premium varying by age, as would be the case for term insurance. ...

3. *The concepts of trust and delegation.* In the absence of ideal insurance, there arise institutions which offer some sort of substitute guarantees. Under ideal insurance the patient would actually have no concern with the informational inequality between himself and the physician, since he would only be paying by results anyway, and his utility position would in fact be thoroughly guaranteed. In its absence he wants to have some guarantee that at least the physician is using his knowledge to the best advantage. This leads to the setting up of a relationship of trust and confidence, one which the physician has a social obligation to live up to. Since the patient does not, at least in his belief, know as much as the physician, he cannot completely enforce standards of care. In part, he replaces direct observation by generalized belief in the ability of the physician. To put it another way, the social obligation for best practice is part of the commodity the physician sells, even though it is a part that is not subject to thorough inspection by the buyer.

One consequence of such trust relations is that the physician cannot act, or at least appear to act, as if he is maximizing his income at every moment of time. As a signal to the buyer of his intentions to act as thoroughly in the buyer's behalf as possible, the physician avoids the obvious stigmata of profit-maximizing. Purely arms-length bargaining behavior would be incompatible, not logically, but surely psychologically, with the trust relations. From these special relations come the various forms of ethical behavior discussed above, and so also, I suggest, the relative unimportance of profit-making in hospitals. The very word, "profit," is a signal that denies the trust relations.

Price discrimination and its extreme, free treatment for the indigent, also follow. If the obligation of the physician is understood to be first of all to the welfare of the patient, then in particular it takes precedence over financial difficulties.

As a second consequence of informational inequality between physician and patient and the lack of insurance of a suitable type, the patient must delegate to the physician much of his freedom of choice. He does not have the knowledge to make decisions on treatment, referral, or hospitalization. To justify this delegation, the physician finds himself somewhat limited, just as any agent would in similar circumstances. The safest course to take to avoid not being a true agent is to give the socially prescribed "best" treatment of the day. Compromise in quality, even for the purpose of saving the patient money, is to risk an imputation of failure to live up to the social bond.

The special trust relation of physicians (and allied occupations, such as priests) extends to third parties so that the certifications of physicians as to illness and injury are accepted as especially reliable (see Section II.B above). The social value to all concerned of such presumptively reliable sources of information is obvious.

Notice the general principle here. Because there are barriers to the information flow and because there is no market in which the risks involved can be insured, coordination of purchase and sales must take place through convergent expectations, but these are greatly assisted by having clear and prominent signals, and these, in turn, force patterns of behavior which are not in themselves logical necessities for optimality.

4. *Licensing and educational standards.* Delegation and trust are the social institutions designed to obviate the problem of informational inequality. The general uncertainty about the prospects of medical treatment is socially handled by rigid entry requirements. These are designed to reduce the uncertainty in the mind of the consumer as to the quality of product insofar as this is possible. I think this explanation, which is perhaps the naive one, is much more tenable than any idea of a monopoly seeking to increase incomes. No doubt restriction on entry is desirable from the point of view of the existing physicians, but the public pressure needed to achieve the restriction must come from deeper causes.

The social demand for guaranteed quality can be met in more than one way, however. At least three attitudes can be taken by the state or other social institutions toward entry into an occupation or toward the production of commodities in general; examples of all three types exist. (1) The occupation can be licensed, nonqualified entrants being simply excluded. The licensing may be more complex than it is in medicine; individuals could be licensed for some, but not all, medical activities, for example. Indeed, the present all-or-none approach could

be criticized as being insufficient with regard to complicated specialist treatment, as well as excessive with regard to minor medical skills. Graded licensing may, however, be much harder to enforce. Controls could be exercised analogous to those for foods; they can be excluded as being dangerous, or they can be permitted for animals but not for humans. (2) The state or other agency can certify or label, without compulsory exclusion. The category of Certified Psychologist is now under active discussion; canned goods are graded. Certification can be done by nongovernmental agencies, as in the medical-board examinations for specialists. (3) Nothing at all may be done; consumers make their own choices.

The choice among these alternatives in any given case depends on the degree of difficulty consumers have in making the choice unaided, and on the consequences of errors of judgment. It is the general social consensus, clearly, that the *laissez-faire* solution for medicine is intolerable. The certification proposal never seems to have been discussed seriously. It is beyond the scope of this paper to discuss these proposals in detail. I wish simply to point out that they should be judged in terms of the ability to relieve the uncertainty of the patient in regard to the quality of the commodity he is purchasing, and that entry restrictions are the consequences of an apparent inability to devise a system in which the risks of gaps in medical knowledge and skill are borne primarily by the patient, not the physician.

Postscript

I wish to repeat here what has been suggested above in several places: that the failure of the market to insure against uncertainties has created many social institutions in which the usual assumptions of the market are to some extent contradicted. The medical profession is only one example, though in many respects an extreme one. All professions share some of the same properties. The economic importance of personal and especially family relationships, though declining, is by no means trivial in the most advanced economies; it is based on non-market relations that create guarantees of behavior which would otherwise be afflicted with excessive uncertainty. Many other examples can be given. The logic and limitations of ideal competitive behavior under uncertainty force us to recognize the incomplete description of reality supplied by the impersonal price system.