The Not-For-Profit Hospital as a Physicians’ Cooperative

By Mark Pauly and Michael Redisch*

The private, nonprofit hospital has usually been regarded by economists as an organizational anomaly. In particular, it has been alleged that the predominance of the not-for-profit structure within the American hospital system is associated with a weakening of the usual market constraints of competition and profit orientation. As a result, this critical element in any analysis of the medical care system in the United States has usually been modeled with a mixture of anecdote and ad hoc assumption. It is typically assumed that “all objectives of nonprofit organizations can be described in terms of some type(s) of output (broadly defined) or capital stock.”1 William Baumol and Howard Bowen describe these goals as “bottomless receptacles into which limitless funds can be poured” (p. 497).

Model variation occurs as investigators place combinations of key variables in either the objective function or the constraint set of the hospital. Joseph Newhouse (1970) and Martin Feldstein (1971) studied the implications of the maximization of quantity-quality subject to a budget constraint. Millard Long’s model is one of quantity maximization subject to both a budget and a quality constraint. Paul Ginsburg assumed maximization of weighted output subject to a budget and an availability of capital constraint. Maw Lin Lee included types of physical capital in the hospital objective function. And Melvin Reder talked of hospitals trying to maximize “the weighted number of patients treated (per time period), the ‘weights’ being the professional prestige to the doctors attending them” (p. 480).

This last model is the only one to (even) hint at a nonpassive role for the physician in a model of hospital behavior.2 In this paper we propose an alternative model in which the physician emerges as a traditional income maximizing economic agent who is “discovered” in a decision-making role within this not-for-profit enterprise. Our model is similar to the model of the firm customarily used by economists, in that it is based on the assumption of net income maximization. Only a somewhat unusual definition of net income is needed to enable us to apply in our short-run analysis many of the conclusions of the orthodox model of the firm. In the longer run, however, our model, while still based on net-income maximization, yields different predictions about the institution’s response to changes in demand and supply parameters. Furthermore, it may be possible to generalize parts of our model to other private, nonprofit service firms such as universities and symphony or-

---

* Northwestern University and the Office of Health Policy Analysis and Research, Department of Health, Education, and Welfare, respectively. Research support was provided by the Health Services Research Center of Northwestern University and the American Hospital Association and the Health Economics Research Center of the University of Iowa. Portions of this paper were presented at the Health Economics Session of Allied Social Sciences Meetings, December 1969. Helpful comments were received from R. Tollison, C. M. Lindsay, and James Buchanan. Views expressed are those of the authors.

1 See Paul Ginsburg, p. 42.

2 Paul Feldstein and Carl Stevens have discussed the role of the physician in the hospital, but have not provided an explicit model.
orchestras, which produce and sell services to individual consumers, many of whom are not poor.

Specifically, we assume that the group of attending physicians on the hospital’s staff enjoys de facto control of the hospital at any point in time. Given this assumption, we develop a model in which the hospital operates in such a way as to maximize the net income per member of the physician staff. Results are obtained which are similar to those derived from models of producers’ cooperatives in Yugoslavia and collective farms in the USSR. The physician plays a role analogous to that of the Yugoslav worker and the Russian peasant. Our results are also similar to those obtained from the “theory of clubs,” developed by James Buchanan and others.

I. The Model

We simplify the problem initially by assuming that patients pay the full market price for care, and that the decision-making group in the hospital is able to impose its collective will on individual members. The implications of weakening these assumptions to allow for customary forms of health insurance and for imperfect cooperation among controlling individuals will be discussed in later sections.

The product produced in the hospital is hospitalization services. We shall assume that this output can be represented by a single variable $Q$. To produce this output, physical capital ($K$), nonphysician labor ($L$), and physician or medical staff labor ($M$) are used. The production process can be summarized by the production function

$$Q = F(K, L, M)$$

In European countries in which physicians who treat patients in the hospital are employed by and paid by the hospital, this three-input production function is the obvious one. But in the United States, the hospital patient is subject to two separate billings. The hospital charges him only for the use of capital and nonphysician labor. The physician presents a separate bill for the use of his “personal” physician’s services. This dual billing system has led to a conceptually false dichotomy in much of the health economics literature. The physician and hospital are often viewed as independent economic entities selling services in functionally segmented health markets. This view appears to provide the rationale for the hospital-administrator-oriented, output-maximization theories of hospital behavior discussed earlier.

We propose an alternative view. It seems obvious that the patient’s demand is primarily for the service produced by the physician and hospital acting in combination, not for the separate components, even though there probably is a separate demand for some attributes, such as amenities, or additional patient days for recuperating, that the hospital alone produces.

The critical assumption of our model is that the physician staff members enjoy de facto control of hospital operations and see to it that hospitalization services are produced in such a way as to maximize their net incomes. The appearance of physician control is not hard to establish. The staff physicians have direct control over the number and types of patients admitted and over the types of services

---

3 Derivatives of two quite different surrogates for hospital output have been used most often by economists doing empirical research. One is based on the number of inpatient days and outpatient visits at the hospital while the other is concerned with the number of cases treated in the hospital. The “case treated” corresponds most closely with the measure of output implied by our model.

4 This was the form used by M. Feldstein (1967) in his study of hospitals in the United Kingdom.
they receive; they control output. The staff physicians can determine, within rather broad limits, what use of the hospital will be made in treating a patient; they control many of the production decisions. They have indirect control over many other aspects of the hospital’s operation, such as capital investment and the level of nursing care, in the sense that no administrator can afford to incur the displeasure of the medical staff, interfere with medical staff prerogatives, or make decisions which will deter large numbers of physicians from remaining on the hospital’s staff or using that hospital for their patients. The trustees, who have nominal control over the hospital’s operation, usually look to the medical staff when making decisions on operations or capital investment.

We first assume that the physicians on the staff of a hospital at any point in time act in such a way as to maximize the sum of the money incomes of all staff members. Such an assumption implies a process of group decision making resulting in a kind of perfect cooperation not likely to be observed in practice. It also ignores nonmonetary components of a physician’s income, such as leisure time and prestige, which are likely to be of some importance. Nevertheless, this model is useful as a benchmark from which to consider the effects of alternative assumptions.

We postulate an economic short-run period as one in which the number of physicians on the hospital staff, $M$, remains constant. Each physician is presumed to supply a constant, homogeneous amount of medical input.

The patient’s demand is primarily for the service produced by the physician and hospital acting in combination, not for the separate components. This can be formalized by postulating a demand function for “hospitalization services” faced by the physician staff that takes the form

$$Q = Q(P_T), \quad \frac{\partial Q}{\partial P_T} < 0$$

where $P_T$ is the combined price paid by the patient for the physician and hospital components.\(^5\)\(^6\)

We also assume that the hospital component of $P_T$ is set so as to allow the hospital to just break even, with no gain or loss.\(^7\) That is, we assume that the hospital price $P_h$ is set to produce the equality:

$$P_hQ = wL + cK$$

where $P_h$ is the unit price the hospital charges for use of nonphysician labor and capital, $w$ is the wage rate for nonphysician labor, and $c$ is the user cost of capital.\(^8\)

---

\(^5\) Some empirical justification of this assumption may be found in Donald Yett et al, where it was estimated that the elasticity of demand for hospital output with respect to surgeons’ fees is 0.7.

\(^6\) If the market for output is perfectly competitive, $|\partial Q/\partial P_T|$ will be infinite; otherwise, the individual hospital demand curve for output will have a negative slope.

\(^7\) If the hospital received contributions, it may set a target loss equal to the contributions, but this will not alter our analysis. Moreover, after the fact the hospital may have a profit or loss, but this is assumed to result wholly from stochastic factors.

\(^8\) The interpretation of the user cost of capital $c$ is worth comment. When capital is provided through borrowed funds, the interpretation is clear; $c$ is equal to $(r+d)P_K$, where $r$ is the interest rate at which the funds were borrowed, $d$ is the depreciation rate, and $P_K$ is the price of capital goods. When unrestricted donations are used to pay for the marginal unit of capital, the user cost is $(r'+d)P_K$, where $r'$ is the opportunity cost of using contributed funds for hospital physical capital, i.e., the rate which could have been earned on those funds if they had been invested elsewhere (say, in government bonds). When donations are made with the restriction that they be used for physical capital investment, they will affect the marginal user cost of capital only if the hospital receives so much in donations that it does not have to turn to any other source for funds for capital investment (unless, of course, the conditions for contribution of restricted funds specify a certain amount of the hospital’s own funds as matching payments). If restricted donations fall short of the amount which, given the interest rate $r$, the hospital wishes to invest so that the hospital borrows, the relevant marginal user cost of capital must involve the interest and depreciation rates. Except for the case in which restricted donations are so large that the amount
The hospital is to be run so as to maximize the net incomes of the physicians on the staff at any point in time. If the number of physicians in the short-run analysis is taken as given at $\bar{M}$, the problem is to maximize $P_rQ - P_sQ$ (which is equal to $P_rQ - \omega L - cK$) subject to the production function (1), with the level of $M$ set at $\bar{M}$, and the demand curve (2). This problem is obviously identical to that facing an orthodox profit-making firm with one input held constant. The marginal conditions for optimal employment of labor and capital are the same, namely, that marginal factor costs equal their respective marginal revenue or value products.  

It may be useful at this stage to contrast the model of the nonprofit hospital just developed with the orthodox model of the profit-maximizing firm. In the latter case, all labor inputs and capital services financed by debt are paid their competitive costs. Nondebt capital then obtains the residual income, which is usually assumed to consist of payment of the opportunity cost of that capital (normal profits) and economic profit. The only difference between this model and the physician-profit maximization model of the hospital is that in the latter it is the physician input, rather than the nondebt capital input, which obtains economic profits, the residual income. If a profit-maximizing firm submitted two bills for its services—one just covering the cost of labor and debt-financed capital, produced in a “nonprofit” firm, and the other from a separate legal institution covering the services of equity capital, the analogy would be complete.

II. Long-Run Individual Hospital Equilibrium

The number of physicians on the staff of any hospital obviously is not fixed, but is variable over time. What determines the size of the hospital’s staff? The answer to this question depends critically on the assumption made about the hospital’s staffing policy. We shall outline the results of three alternative policies—closed staff, open staff, and a policy in which new physicians can be hired by the hospital.

Closed staff

Many hospitals in the United States restrict staffing privileges; they do not permit any physician to join the hospital’s staff just because he wishes to do so, even if he is licensed to practice medicine and surgery. The decision on whether or not to admit a new member to the staff (or whether to replace a member who has left) is made by the existing members of the hospital’s staff of physicians. If we
assume that once a physician is admitted to the staff, he has privileges identical to those of the existing members, the appropriate maximand for the hospital appears to be the maximization of net income per physician, \( Y_M \). Physicians will be willing to add members to the staff as long as it causes each member's net income to rise.

This implies that the objective function to be maximized is

\[
Y_M = \frac{(P_T Q - cK - \omega L)}{M}
\]

subject to the production function (1) and the demand curve (2). The necessary first-order conditions for an extremum become

\[
\begin{align*}
\omega &= P_T \frac{\partial Q}{\partial L} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial L} Q \\
c &= P_T \frac{\partial Q}{\partial K} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial K} Q \\
Y_M &= P_T \frac{\partial Q}{\partial M} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial M} Q
\end{align*}
\]

In long- or short-run equilibrium, the physician-hospital conglomerate firm that we have postulated will equate the marginal supply price of all nonphysician inputs to their respective marginal revenue or value products. However, in our model, physicians all share equally in the residual income of this health enterprise, the shares depending on their assumed equal shares of a total output. Condition (5c) states that physician staff size is determined in long-run equilibrium by equating the marginal revenue product of physicians to the net average revenue product of the physician staff. This makes intuitive sense. The hospital "pays" for new physicians by allowing them a proportionate share of total output and, hence, of net revenues. Staff physicians will want to welcome warmly a new member as long as his contribution to total revenues of all staff physicians is greater than the average current income per physician which he receives.

Of course, condition (5c) cannot be satisfied unless there are physicians willing to work at the hospital for the earnings available. That is, the equilibrium value of \( Y_M \) must be at least as large as the income stream available to a physician in his next best opportunity. There will be a supply curve of physicians to any hospital. The shape of this curve will depend in part on the income a physician could get in other hospitals, and his valuation of other uses of his time, both as leisure and as office practice.

Figure 1 depicts the long-run equilibrium position of a hospital operating in an urban, physician-intensive environment. The physician supply curve, SS, may therefore be assumed to be approximately infinitely elastic, and we also assume that it is at a low level relative to income possibilities in this particular hospital. Within the hospital, capital and nonphysician labor take on short-run optimal values as physician staff size, \( M \), varies along the horizontal axis. \( ABC \) thus represents the maximum attainable income per physician for each specific value of \( M \). Returns to scale and elastic demand lead initially to the upward sloping segment of \( ABC \), but eventually decreasing returns and diminishing marginal revenue cause the curve to turn down. The maximum maximorum, \( Y_M^* \), of this set of short-run maximums is reached at the intersection of the marginal revenue product and net average revenue product curve, when physician staff size reaches its long-run equilibrium value of \( M^* \).

This model is very similar to those developed by Benjamin Ward (1958, 1970), Evsey Domar, Walter Oi and Elizabeth Clayton, Jaroslav Vanek, and others who explain the economic behavior of the Soviet collective farm or the Yugoslav producers' cooperative. The physician
plays a role analogous to that of the Russian peasant and the Yugoslav worker. Institutionally, the arrangements differ because the physician receives his "share" of the enterprise's income directly from the sale of his output, whereas the worker on a collective receives it from a common pool. But this is because, in the case of the hospital, output can usually be directly assigned to particular staff members.

Our model of the hospital shares certain of the seemingly paradoxical conclusions of these cooperative-collective models. Supply response to changes in product and factor market conditions can be in perverse directions. An upward shift in the demand curve for hospital output could result in a new equilibrium with higher price levels, lower output, and fewer physician staff members to share the greater total net revenues of the physician-hospital conglomerate enterprise. "Members have an incentive to contract membership to hoard the spoils" of a demand increase (see Oi and Clayton, p. 43). An increase in a factor price may lead to an expansion of operations to help spread the misery around among a larger group of individuals. On the other hand, a lump sum subsidy, such as a philanthropic contribution, will decrease output and staff size. "Even when the co-op moves in the same direction as a capitalist firm, its response is usually more sluggish. For market stability, the picture is not particularly reassuring" (see Domar, p. 739).

The reason for this result can be sketched out briefly. Suppose a hospital faces a given price for output, and price increases by some amount. This will produce a proportionate increase in the marginal revenue product of physicians, assuming that $K$ and $L$ are held constant, but the income per physician ($Y_M$) will increase more than proportionately. The

---

10 Unambiguous results concerning the direction of supply response cannot be determined unless specific restrictions are placed on the nature of the demand shift and the form of the production function.

11 For a more extensive development, see the references cited in the preceding paragraph.
maximum average income would be attained at a smaller value of $M$. The values of $K$ and $L$ will not, of course, remain constant. They will increase in response to increases in their marginal products. But $M^*$ will still decline unless increases in $K$ and $L$ increase the marginal revenue product of physicians and do so by a large enough amount to offset the increase in average income. This need not happen; $Q$ and $M^*$ may decrease, or not increase much, whereas in a profit-seeking firm the use of an input and the amount of output would almost certainly increase with a rise in the price of output.

*Discriminatory Sharing or Hiring Model*

To consider the consequences of altering the “equal sharing” assumption, we suppose that the hospital depicted in Figure 1 is allowed to hire physicians at the supply price, $OS$. The hospital will then organize production in the same way as would a profit-maximizing firm. The physician input will be increased until the marginal revenue product and the marginal supply price of physicians’ services are equal to each other. Staff physicians will be able to capture, in their own incomes, the excess of the marginal products of the intra-marginal hired physicians over their supply price. Equilibrium will be at $MD$. But note that there is explicit discrimination in returns to homogeneous labor in this situation; the ability to sustain a stable equilibrium under these circumstances is highly suspect. As Pauly has shown, a system of clubs in which some identical persons receive less than others is not likely to be stable. It may well be that the internship-residency programs so prevalent in the United States may be an institutionalized method of getting around the inherent instability in the discriminatory hiring model by creating artificial, functionally viable distinctions among homogeneous physicians. Determination of the ratio of “partner” physicians to “hired” physicians is likely to be arbitrary. The economic well-being of those physicians left with full staff privileges varies inversely with this ratio.

*Open Staff Model*

Alternatively, we can retain the “equal shares” assumption, and examine the economic behavior of hospitals that do not restrict entry to their physician staff. Any licensed physician who chooses to do so may become a “full partner” in the firm. Equilibrium at $M_e$ in Figure 1 is characterized by equality of average income per member of the physician staff and the marginal supply price of physicians services.

Of course, for a hospital in a rural area with few physicians, or in an area where many physicians have attractive alternatives to membership on that hospital’s staff, the situation might be somewhat different. Faced with a sharply rising supply curve for physicians’ services, $S'S'$ in Figure 2, the physician staff will be in equilibrium at $M_{ee}$. The hospital might as well call its policy “open staffing,” since it would be willing to add new members in order to move up the rising part of the net average product curve and increase income per staff member. Such a “frustrated closed staff” hospital only needs to adopt a closed staff policy when the number of staff members reaches $M^*$.

Note that the discriminatory-hiring and open-staff hospitals, either the “true”

---

12 Interns and residents serving in U.S. hospitals currently comprise about 15 percent of all MD’s and, of course, provide a considerably higher percentage of hospital based physicians’ services. In several of the cost function studies, there have been attempts to estimate the effect these physician trainees have on hospital costs and revenues. Our model would suggest that it might prove more fruitful to analyze the ways in which trainees can increase the incomes of those physicians with regular staff privileges.
open-staff hospital or the frustrated closed-staff one, do not exhibit the potentially perverse supply responses that were uncovered in the closed staff model. Increases in demand will lead to increases in physician staff and output.

III. Long-Run Industry Equilibrium

The closed-staff hospital in the long run tries to adjust its physician staff size to achieve maximum income per physician. But if physicians in a hospital achieve incomes below those of other identical physicians on the staffs of other hospitals in that area or in other areas, the low-income physicians may wish to join the staffs of high physician-income hospitals. If they are prevented from joining the staffs of existing hospitals, as would occur if those hospitals were large enough so that average income was maximized, it will be worthwhile for $M^*$ of those rejected physicians to join a new hospital, a duplicate in size of the old one. Indeed, the formation of new hospitals will continue as long as higher incomes can be earned, bidding up the supply price of physicians and bidding down the price of final output (and hence the curve of average income) until a position of long-run equilibrium is attained at which average income, marginal income, and marginal supply price are equal to each other and equal across hospitals. In the open-staff model, long-run equilibrium is also reached by the formation of new hospitals, but here the new hospitals draw off excess members from existing hospitals, raising average incomes since average size shrinks. This process will continue until formation of a new hospital does not raise average income. Finally, in the discriminatory-hiring model, physicians who are paid less than staff physicians (i.e., hired) will find it advantageous to try to join a hospital of their own in which they are all staff members and receive incomes equal to those of physicians in existing hospitals.

Thus these models have identical long-run industry equilibria (when the number of firms is variable), but the closed-staff model has a radically different long-run firm equilibrium, and may move in a perverse direction from that of the discriminatory-hiring firm. The open-staff model, on
the other hand, responds in the same way as would the discriminatory-hiring (or profit maximizing) hospital, but is likely to have a quantitatively different response. Although the final industry outcomes under these institutional arrangements are the same, the adjustment process by which that outcome is reached is very different in each case.

IV. Imperfect Cooperation on the Hospital Staff

In this section we wish to relax the assumption of perfect cooperation by the hospital’s staff of physicians. The individual physician may in fact have direct control over the process of producing output. He is able to order use of nursing and other inputs for his patients, in his character of professional expert. At best, “the hospital” can determine only a stock of inputs; the physician controls the flow of services from them. This is in contrast to a producers’ cooperative (or even a nonprofit university) in which, presumably, a central management is somewhat more able to prevent individualistic behavior. Moreover, each physician’s income from the hospital-connected services depends entirely on his own output, in contrast to the producers’ cooperative model in which an individual worker’s income partly depends on his own output but also depends on his share of total profits or surplus income. The physician shares group income insofar as he shares group output.

We assume, as before, that the hospital price is set at a level which permits the hospital to break even. It is now easier to see why this assumption is itself a consequence of the physician profit-maximization hypothesis. Other things being equal, hospital profits mean higher hospital prices, which reduce physician incomes. To be sure, hospital profits could be used for capital investment which may enhance future physician incomes, but 1) physicians may not want to invest in the hospital—and 2), as Eirik Furubotn and Svetozar Pejovich have noted, the fact that property rights are not vested in investment from profits makes it even less desirable than owned investment.

Since prices for services are not (yet) set at marginal cost,\textsuperscript{13} and since the physician does have the power to direct the application of labor and capital services in producing his output, noncooperative behavior can occur. Suppose the hospital follows the policy of charging an all-inclusive daily rate. Suppose ordering the use of more labor in the production of some output will permit the physician to charge more for that unit of output, because the “quality” of the output is enhanced. If the application of more nonphysician labor raises the price \( P^t_i \) that can be charged by the \( i \)th physician but if the same price \( P^t_H \) is charged for all units of hospital output, regardless of whether or not extra labor is ordered, the physician will want to use labor for his \( Q_i \) patients up to the point at which the following equality is satisfied:

\[
\frac{\partial P^t_i}{\partial L} Q_i = \frac{Q_i}{Q} w
\]

That is, the individual physician only considers a fraction \( Q_i/Q \) of the cost of the labor he orders to be employed, whereas group profit maximization would require him to take account of the costs his action imposes on other physicians’ patients. Consequently, imperfect cooperation leads

\textsuperscript{13} In fact, it does not appear that hospitals do price at marginal cost. They have moved away from the use of an all-inclusive daily rate, the same for all patients, toward itemized bills, but they are still far from charging marginal cost as a price. And in theory, it is unlikely that prices could be set equal to marginal cost. Unless marginal cost just equals average cost, marginal cost pricing by the hospital of its services would violate the no-profit, no-loss constraint assumed above. Moreover, the use of different prices for different qualities of outputs involves transactions and information costs, which may be substantial.
to the use of too much labor relative to the amount that would be used if group incomes were maximized with perfect cooperation.

How great these departures from perfectly cooperative usage will be depends upon how well the staff as a group can control the actions of its individual members. It seems reasonable to suppose that the magnitude of the departure from cooperative behavior should get larger as the size of the staff gets larger. There are at least three reasons for making this assumption. 1) When the staff size is small, each physician bears a larger share of the cost of his own actions. 2) When the staff size is small, departures from cooperative behavior on the parts of others are more noticeable to any single member. 3) When the staff size is small, mutually agreeable group decisions are more likely to be arrived at, since the cost of decision making is less.

In terms of the geometric analysis, introduction of imperfect cooperation in this way shifts the curve of net revenue per member down and to the left. The conclusion is obvious; lack of perfect cooperation makes the (second best) optimal size of the staff and hospital smaller than it would otherwise be.

V. The Effect of Insurance

To this point we have assumed that patients confronted the full market price for care. Insurance coverage of medical costs may weaken the applicability of this assumption. The hospital is typically paid on the basis of costs incurred. If there are no copayments or deductibles at all, this arrangement effectively eliminates market control over the hospital component of the price of hospitalization, since higher hospital costs and prices have no effect on the cost actually paid by the patient (except through the premium, which only matters if the premium rises so high that the person drops coverage). Moreover, the allocation of costs within the hospital to insured patients, especially Blue Cross members, is often done on average cost basis. The only restraint on the physician’s prescribing the use of hospital capital and labor for his fully insured patient is the upward pressure his behavior would exert on the prices paid by his noninsured patients. Conversely, under average cost allocation some of the market discipline is lost for the noninsured by the transfer of some costs to the insured.

If every patient had full coverage cost-based hospital insurance, there would be no constraint on the amount of capital and labor that physicians combine with their services. Capital and labor would therefore be employed up to the point at which the marginal contribution of each to the physician’s revenue was zero. This produces “Cadillac-quality” medicine. The only constraint on the use of these inputs would be offered by the upper limit on the number of things a hospital can do for a patient which might have some justification. Over time, technological change might be expected to relax even this constraint.

When insurance covers part of each patients’ hospital bill, the factor prices of hospital inputs K and L are effectively reduced as far as the group of physicians is concerned. One would expect an increase in the usage of hospital inputs relative to physician inputs for producing a given output. Hospital unit costs would rise. Thus our model provides an explanation of the positive relationship between hospital insurance coverage and hospital unit prices and costs found by M. Feldstein (1971). In addition, our model yields the important result that, ceteris paribus, increased

---

14 A similar analysis of the effect of size is provided by Mancur Olson. Other applications to medical care can be found in the work of Newhouse (1972) and Mather Lindsay.
hospitalization insurance should increase physician prices and physician incomes.

VI. Toward a Theory of the Not-For-Profit Enterprise

To this point we have taken the not-for-profit nature of the typical American hospital as given. In this section we offer an explanation of why this organizational form has arisen.

There are two ways in which the hospital might operate if it were on a profit-maximizing basis. It might combine the services of nonphysician labor and physical capital, and sell them to or through the physician, who combines them with his own input to produce the output of hospitalization services. Alternatively, the hospital might perform the job of combination itself, hiring the physician input and selling the final output. In either case the direct control over the use of nonphysician labor and capital would not be held by the physician. In the first case, he would have to use the market for control, which is not always efficient in the sense of minimizing all costs, including transactions costs, as direct control. In the second case, he himself would be under the direction of the supplier of equity capital. It is surely possible that there are some products which are not produced efficiently when a representative of the owners of equity capital directs their production, or when the market is used to organize the production process instead of the use of direct controls within a single organization. The most efficient method might be for the supplier of an important component of the labor input to direct the production process.

This would tend to happen when human capital is important in the production of some output and when the flow of services of that human capital cannot well be directed from outside, but is controlled by the person in whom the capital is embodied. As Armen Alchian and Harold Demsetz have noted, the wage system tends to break down when marginal products cannot be monitored closely. This may be a reasonable conjecture in the case of the production of hospital services. Many of the decisions the physician has to make are decisions which cannot be supervised directly, and which have contingent outcomes. There probably needs to be some incentives for the physician. Financial interest in the outcome of his actions is one such incentive, and that incentive is at its greatest when the physician bears the full residual income, when the consequences of his actions are not spread over suppliers of physical capital.15

The production process requires some physical capital. In a labor-managed firm in which most of the assets of labor are embodied in nontransferable human capital, not all of the physical capital can be borrowed, since collateral cannot be provided. Another necessary condition therefore for the emergence of the not-for-profit form would seem to be the willingness of individuals to contribute for its equity capital. In principle, contributions could either be voluntary or provided through government. Where voluntary contributions are sufficient, government contributions would not be expected to emerge. On the other hand, voluntary contributions may arise in precisely those cases in which the government fails to act. They may also arise in cases in which the government through tax deductibility

15 To see this, think of each physician as a "firm." The socially most efficient institutional arrangement is the one which maximizes the net present value of this firm. The present value is a contingency, depending on the state of nature (for example, what's really wrong with his patients, whether an epidemic occurs, etc.) and the amount of effort that the physician makes. The amount of effort, in turn, depends upon the share of profits the physician receives. The appropriate share for the physician, even given the greater risk he bears, may be approximately equal to one.
subsidizes private contributions. The source of contributions, whether unsubsidized voluntary, subsidized voluntary, or governmental, is not critical to the argument, except to the extent that one form (for example, government) implies more external control over physicians' actions than another.

These contributions could be motivated by a desire on the part of contributors to make output available to themselves or to those whom they would like to see consume it. That is, the motivation could either be based on the potential receipt of private benefits or of external benefits. Contributions are a logical way for potential purchasers of the outputs of labor-managed firms to make possible production of the output, which they or those about whom they are concerned will use. If there are barriers to entry by profit-seeking firms (as there are in higher education and, to some extent, in the hospital industry as well), potential consumers may be willing to contribute if that is the only way that output, which yields them consumers' surplus, can be made available. It is not surprising that private not-for-profit firms which sell output—hospitals, universities, symphony orchestras—tend to provide output which is used not by the poor, but partly by the contributors themselves.  

VII. Conclusion

The main thrust of the model we have suggested here, and the one which differentiates it from models of the not-for-profit hospital that have been suggested by others, is the use of the maximization of physicians' income as the characteristic function. The potential absence of perfect cooperation distinguishes it from similar models of producers' cooperatives. In a methodological sense, our model seems to be more attractive than those which simply assume that the not-for-profit organization maximizes a variable such as "quantity of output," because it explains what the organization does in terms of the economic motivation of those who control it.

More importantly, it appears to provide an appealing explanation of some peculiar characteristics of not-for-profit hospitals. The supposed quality consciousness of such hospitals, for example, is easily explained; "quality" is a synonym for application of nonphysician labor and capital in physician-income-enhancing ways, and noncooperative behavior could easily lead to "too high" quality. "Duplication of facilities" probably owes its existence to closed staffing and lack of perfect cooperation. Other aspects of hospital behavior could also be explained by considering their effect on physicians' income; the pattern of investment, for example, might be best explained by changes in the ability of capital to enhance physicians' incomes. The inelastic supply response of hospitals to Medicare and Medicaid is also consistent with our model. Even the average size of hospitals, which seems, by most accounts, to be below the optimal or cost-minimizing level, can easily be explained. In the first place, empirically observed cost curves may be misleading, if we add the physician input. But more importantly, in a period of rising prices our model shows that hospitals will tend to be small, and for two reasons. First, smallness tends to permit maximization of net income per physician. Second, smallness is necessary to permit coordination of the medical staff.  

---

16 This last point is an important consideration. It is sometimes alleged that these firms have attained a nonprofit status so that they may better provide services to the poor. However, the recent experience in this country is for the poor to receive health services from government operated hospitals, to receive education in government operated institutions, and not to partake at all of the output of symphony orchestras, theatre groups, or private universities.
A narrower range of possible observations is consistent with our model than with the general utility-maximization model. Appropriate choice of the variables to enter the utility function can make almost any observed behavior consistent with utility maximization. In particular, the definition of quality is not clear. Our model specifies the variable in the objective function. In principle it will also predict quantitative as well as qualitative responses, in the sense that physician income can be measured while utility cannot. Unfortunately, at present hospital-specific data on physicians’ incomes or prices do not exist which would permit us to provide a conclusive test of the model. Nevertheless, we hope that more data can be made available to test this model.

REFERENCES


