HOSPITAL RATE REVIEW
A Theory and an Empirical Review*

Michael A. MORRISEY
Hospital Research and Educational Trust, American Hospital Association, Chicago, IL 60611, USA

Douglas A. CONRAD
University of Washington, Seattle, WA 98195, USA

Stephen M. SHORTFEL.I.
Northwestern University, Evanston, IL 60201, USA

Karen S. COOK
University of Washington, Seattle, WA 98195, USA

Received September 1982, final version received July 1983

This paper presents a theory of the effects of rate review on hospital operations and organization. Its purpose is to explain the way in which hospitals have responded to regulation. In the development of this theory, the hospital product was viewed as a bundle of services; rate review was looked upon as a ceiling on the value of the bundle. The ceiling creates an incentive to remove elements from the bundle, i.e., to reduce 'quality'. When quality is variable, the effect on utilization becomes indeterminate. The model argues, among other things, that the hospital will change its service complement and its contractual arrangements with physicians and other hospitals. An extension of the organizational theory literature leads to implications concerning the ordering of hospital responses to regulation. The growing body of empirical literature on the effects of hospital rate review is used as an initial test of the major thrusts of the theory. A suggested agenda for further empirical work also is presented.

1. Introduction

This paper presents a theory of the effects of rate review on hospital operations and organization. Its purpose is to explain the way in which hospitals have responded to regulation. The underlying framework used in

*An earlier version of this paper was presented at the Allied Social Science Associations' Annual Meeting, Washington, DC, December 19, 1981. We are indebted to Yoram Barzel and Keith Leffler for useful discussions and to Joseph Newhouse, Joel May, Bruce Steinwald and an anonymous referee for their comments on earlier drafts; the remaining errors are ours alone. The study was funded, in part, by grant no. 04472 from the National Center for Health Services Research. The opinions and conclusions expressed in this paper are solely our own and should not be construed as necessarily reflecting those of our organizations.
the analysis is Barzel’s (1976) application of Lancaster’s (1966) characteristics model. Barzel used this model to demonstrate that ad valorem and per unit taxes have different allocative effects because they provide different incentives for changes in the characteristics comprising the transacted commodity. An ad valorem tax is a tax on the value of the attributes bundle; the incentive created is to remove attributes and sell them tax free. A per unit tax literally applies only to those attributes specified by the tax. As a result, the characteristics are reconfigured to reduce the contribution of the taxed attribute. In Barzel’s example, the durability of the taxed light bulb is increased. In general, the incentive is to minimize the sum of the financial loss caused by the tax and inefficiency loss resulting from the use of a non-cost minimizing bundle of attributes.

In this analysis, we argue that the various services that a hospital provides constitute the bundle of characteristics or attributes that the hospital offers. Unlike earlier studies, which define the patient day or the admission as the hospital product, we view these as elements in the bundle of hospital care; also present are bed rest, special unit care, specified therapies, tests and lab work-ups, among a host of other components. Rate review is viewed as a ceiling on the value of the hospital care bundle and, in the absence of other constraints, it provides incentives to remove services from the hospital care bundle.

The theory yields testable hypotheses concerning case mix, length of stay, and service offerings. A relatively minor extension of the literature on the nature of the firm leads to hypotheses concerning hospital vertical integration as well as contracting among hospitals and between hospitals and physicians. An extension of the organizational theory literature leads to implications concerning the ordering of hospital responses to regulation. Using the theory as a benchmark, the prior empirical research on the effects of rate review on hospital costs, revenues, operations and organization is reviewed. While much of this research is in its infancy, it is consistent with the theory. Finally, a proposed agenda for further research is presented.

2. A theory of rate review

In this analysis, the consumer–patient is viewed as having a utility function consisting of health and other goods. However, health is not purchased directly; it is produced as a function of the consumer’s initial endowment of health and the medical care services he purchases.

Following Pauly (1980), this analysis views the physician as a net income maximizing provider of medical care. Consumers are assumed to be indifferent to the mix of inputs the agent–physician employs. In our model the physician is, therefore, a general contractor employing hospital services and his own time and skill to produce medical care. As a result, consumer–
patients display a demand for medical services that is a function of the full net price of the bundle of medical care; however, they are indifferent to the prices of the various components. Under these conditions it is easy to show that the physician will minimize the cost of producing whatever quantity of medical care he does provide.\(^1\)

The interest of this analysis is in the services provided by the hospital to the physician. The non-profit hospital is viewed as maximizing an objective function consisting of one argument: medical staff net income. This objective function is consistent with the work of Pauly and Redisch (1973) and Pauly (1980) and, as a sign preserving function, necessarily implies that the hospital will employ its inputs to minimize the cost of hospital services provided to physicians.\(^2\)\(^3\) Similarly, the for-profit hospital is assumed to minimize input costs in the process of maximizing the net present value of its operation, i.e., in the for-profit case the residual claimant to the hospital’s net income is the shareholder, not the physician staff.

To begin, we assume the hospital market is price competitive. To further simplify the exposition, it is assumed that physicians have identical concave net income functions and that hospital services can be characterized completely by the level of two attributes, \(X\) and \(Z\). Adopting a model developed by Leffler (1982), units of output \((X)\) are explicitly priced in the hospital market, whereas the vector of attributes subsumed in \(Z\) is not. Of course, the market price of \(X\) is conditional upon the level of \(Z\). We refer to the level of \(Z\) relative to \(X\) as the ‘quality’ per unit of output \((q = Z/X)\). The product transformation function for \(X\) and \(Z\) is assumed to be homothetic with respect to (non-physician) hospital capital and labor inputs, as well as the physician’s own input, which allows us to focus on substitution effects among the factor inputs. Similarly, it is assumed that consumers’ marginal rates of substitution between \(X\) and \(Z\) are independent of income, which permits us to focus on substitution in consumption between \(X\) and \(Z\). Finally, we posit that priced and unpriced attributes are packaged together because of economies of joint production.

Let \(TV = t(X,Z)\) and \(TC = c(X,Z)\) denote, respectively, the total valuation and the total cost functions for hospital care. Assuming positive but

---

\(^1\)Pauly shows this conclusion to be consistent with various forms of insurance; the usual conclusion about overuse of the insured service, relative to the no-insurance model, also is obtained [Pauly (1980, pp. 1–64)].

\(^2\)This implies that the shadow prices of hospital services reflect their marginal costs, given any hospitalization insurance subsidy. It also implies the absence of hospital service cross-subsidization. This construction is analogous to the vertically integrated firm in which intra-firm units set marginal cost transfer prices [Mc Kee and Bassett (1976)].

\(^3\)This construction also implicitly assumes away the allocation of the hospital residual across members of the medical staff. Harris (1977) and Pauly (1980) discuss the implications of the medical staff coordination-free rider problem. In general, the costs of organizing imply a ‘non-optimal’ service bundle, smaller levels of hospital output and medical staff size, and a smaller residual.
decreasing marginal valuation with respect to \( X \) and \( Z \) and positive but increasing marginal costs, market equilibrium is determined by the level of quantity and quality per unit of quantity, which maximize the net incomes of physicians \( (TV-TC) \).

Algebraically, the firm's decision problem can be stated as follows: maximize

\[
P_T X - wL - cK,
\]

such that

\[
g(Z, X) = h(L, K, M),
\]

and given

\[
X = X(P_T, Z),
\]

where

- \( P_T \) = total money price of hospital care,
- \( w, c \) = factor prices for non-physician hospital labor \( (L) \) and capital \( (K) \), respectively, and
- \( M \) = level of physicians' input to hospital care.

2.1. Market equilibrium

Fig. 1 shows the locus of potential competitive market equilibria in \( X \) and price of \( X \) space. The \( A^*A^* \) curve reflects the locus of physician net income maximizing levels of quantity for each level of quality. The coordinates \( P_e, q_e \) and \( X_e \) denote the preconstraint equilibrium price, quality and quantity levels, i.e., the global solution to maximizing physician net income. The quality-quantity mix yielding maximum quantity is found where quantity equals \( X_1 \). The negatively sloped portion of the \( A^*A^* \) curve denotes that region in which the marginal valuation of quantity due to an increase in quality is less than the corresponding marginal cost. Conversely, the positively sloped portion is the region in which upward shifts in the marginal valuation of quantity curve due to an increase in quality exceed the induced upward shifts in marginal costs. The negatively sloped portion, associated with higher quality levels, reflects rapidly diminishing returns to quality increases. For example, the costs of monitoring a hospital production process

\footnote{Net income maximization by physicians and by shareholders implies that hospital resources flow to their highest valued use. At the margin the return to the residual claimant in for-profit and not-for-profit hospitals will be equal as physicians-qua-producers search for cost minimizing hospital services.}
of very high technical quality (and, therefore, likely involving high resource intensity and subtle variations in clinical art) may be expected to be quite high relative to the physician’s valuation of those subtle improvements in technical efficiency.

We posit that rate review acts as a ceiling on average revenues per unit of output and that, in the absence of requirements specifying the unpriced attributes of each unit of output, rate review programs will induce a variety of responses on quality and quantity margins. When a rate review constraint is introduced, the model yields an ambiguous prediction of the quantity response but an unambiguous reduction in quality. Fig. 1 also demonstrates this. As drawn, the example shows a physician net-income maximizing equilibrium at $X_e$, $P_e$, with quality $q_e$. The introduction of a binding rate review constraint defining the hospital product only in terms of the priced characteristics yields a price ceiling $P$, which is less than $P_e$. Holding quality
constant, a hospital would reduce quantity to $X_2$. However, because this is not an efficient solution, net-income can be increased by altering the composition of the hospital service bundle, which is accomplished by reducing quality. The result is lower marginal valuation and marginal cost curves. The process continues until the quantity-quality mix is again on the $A^*A^*$ curve. In short, any move down the $A^*A^*$ curve reduces quality. The quantity effect depends upon the original location on the $A^*A^*$ curve. Had the original market equilibrium been at point $E'$ on the positively sloped portion of $A^*A^*$, the ultimate constrained equilibrium would have involved both lower quantity and lower quality relative to the unconstrained case.

Thus, as fig. 1 makes clear, the direction of the quantity response to the constraint is conditional upon whether the preconstraint equilibrium is in the positively or negatively sloped region of the $A^*A^*$ locus and, for large differences between $P_e$ and $P$, on the magnitude of those differences. While the direction of the quantity response is ultimately an empirical question, certain features of the hospital market serve to establish a reasonable presumption. Within the constructs of the Barzel-Leffler model, an ad valorem subsidy unambiguously increases both quality and quantity. The widespread use of service-benefit style health insurance suggests the presence of such subsidies in the hospital industry and increases the likelihood that the empirically relevant region of potential market equilibria will be in the negatively sloped region, implying increased quantity but decreased quality as the responses to rate review.

2.2. Imperfectly competitive market structures

Our results are not strongly affected by the assumptions of a competitive market structure or hospital maximization of physician net income. To clarify the issues, assume a homogeneously separable consumer utility function, that is, one in which all bundles of $X$ and $Z$ that provide equal services to the consumer also provide equal revenue to the seller [Leffler (1982)].

Consider first the preconstraint equilibrium under hospital monopoly. Physician net income can be maximized by first identifying the quality-quantity bundles that minimize the cost of providing services and then choosing a monopoly level of service. Given a downward sloping service demand function, the monopolist will clearly restrict service relative to the competitive result. However, this may be done by reducing $X$, reducing $Z$, or both. As a result, one can only conclude that both quantity and quality will not be higher under monopoly than under a competitive structure.\(^5\)

\(^5\)It is the case, therefore, that only if the efficient level of quality ($Z/X$) is constant for varying levels of service will the monopoly solution necessarily imply a reduction in output relative to the competitive case.
Now consider the imposition of a binding rate ceiling of the form described above. Holding quality constant, the monopolist would, of course, increase net income by expanding output until the ceiling rate equalled marginal cost (or crossed the demand function, whichever occurred first). The monopolist can do better by then reducing quality. Such an action shifts down marginal costs and shifts in demand but leaves the price ceiling unaffected. Quality would be reduced as long as the extra cost savings, on all units of quantity produced, exceeded the net revenue lost on the units of output given up. As a result, quality declines but quantity is indeterminate, again depending on the starting point.

Alternatively, suppose one assumes that the non-profit hospital maximizes a utility function that includes quantity and quality subject to a minimum profit constraint. In such a case managers are usually characterized as dissipating their monopoly rents by choosing higher quantity and quality than would the perfectly competitive, profit maximizing firm. In that case, the imposition of a binding rate review constraint would lead to a reduction in quality but an increase in quantity. The magnitudes of the responses would, of course, depend on the manager’s utility weights for quantity and quality.  

The comparative static predictions of the model, to this point, do not appear to be particularly sensitive to the Pauly–Redisch assumption of physician net income maximization by non-profit hospitals. A quantity and quality maximization model has been shown to yield similar output and quality attribute effects of rate review regulation, at least for a special case. Similarly, a model of utility maximization over case types [Hornbrook and Rafferty (1982)] generates comparable predictions for the effects of per diem, per admission and per diagnosis-specific admissions. Assumptions regarding the maximand of the hospital appear to principally influence the optimal

---

*To see this in a special case consider the following utility maximization model:

\[
\max U = U(X,q) \quad \text{subject to} \quad P(X,q)X - c(q)X = \Pi_0.
\]

This implies

\[
\max L = U(X,q) - \lambda(P(X,q)X - c(q)X - \Pi_0).
\]

First-order conditions for a maximum (diminishing marginal valuation and rising marginal cost of \(X\) and \(q\), respectively, assure that this represents a maximum):

\[
\frac{\partial L}{\partial X} = 0 = U'_x - \lambda(P'_x X + P - c(q)),
\]

(4)

\[
\frac{\partial L}{\partial q} = 0 = U'_q - \lambda(P'_q X - c'(q)X),
\]

(5)

These equations can be rewritten as

\[
X = U'_x / \lambda(P'_x + P/c(q)/P'_q)
\]

and

\[
X = U'_q / (\lambda(P'_q - c'_q)).
\]

(4')

(5')

Ignoring cross-effects across equations and assuming \(P'_x\) and \(P'_q\) to be constant, eq. (4') implies that as \(P\) is reduced, \(X\) increases. Eq. (5') implies that as \(X\) is increased, \(q\) must be reduced to achieve equilibrium, i.e., reducing \(q\) implies increased \(U'_q\), the marginal utility of quality. Sloan and Steinwald (1980, pp. 20–28) derive negative effects on both quality and quantity using an alternative special case.
levels of the arguments in the objective function, whereas comparative static predictions of changes resulting from changes in exogenous parameters are, for the most part, unaffected.

2.3. Firm equilibrium

Under the assumptions of competitive hospital product markets and cost minimizing firm behavior, the market marginal cost schedules in fig. 1 are obtained from the horizontal summation of the firm's marginal cost schedules above the minimum point on their efficient long-run average cost curves. The firm's short-run response to a binding rate review constraint will be to suffer losses prompting adjustment in efficient quality and quantity. In the long run, firms with relatively higher costs and less flexibility in adjusting their production processes will leave the hospital industry.

2.4. Operational considerations

The model predicts reduced quality, i.e., a reduction in the unconstrained attributes relative to the constrained attributes of the hospital product. Thus, there is an incentive to reduce services per day or per admission. The changed service mix implies a narrowing of the medical staff specialty distribution. Those specialties employing costly but unconstrained services — perhaps nuclear medicine or pediatric cardiology — will be less frequently represented on staff. This, in turn, implies a reduction in the range and complexity of the hospital's case mix.

However, this scenario goes too far too fast. The incentive created by the rate review constraint is to find those means of adjusting the unpriced attributes, i.e., the quality, such that the downward displacement of the marginal valuation and cost curves allows the hospital and its physicians to meet the revenue constraint and attain maximum net income. Literal elimination of services may not satisfy the optimality condition.

Consider an artificial but not wholly unrealistic circumstance in which the rate review authority specifies a maximum revenue per diem constraint. One could eliminate intensive care unit services. An alternative, which also potentially satisfies the constraint but appears to offer higher net benefits, is to expand the duration of hospital stay, thereby reducing the services provided per day. The result is that the bundle of unconstrained attributes is reduced relative to the constrained 'patient day'. Consider a second rate review program that constrains revenue per admission. In this case, duration of stay could be extended by a discharge-readmission scheme without altering services or case mix.

A rate review program does not only provide incentives to eliminate services and to change the duration of a hospital stay, its effects pervade the
operations and organization of the firm. For example, most rate review programs impose no limitations on hospital-based physicians. One means of satisfying the rate review constraint is to remove such physicians, typically anesthesiologists, radiologists, pathologists and emergency department specialists, from the hospital payroll. The physician may continue to perform precisely the same functions in precisely the same fashion as before but now bills the patient separately from the hospital.\footnote{The desirability of this form of unbundling is limited by the extent of insurance coverage for services received in the hospital but billed separately. Further, as we discuss below, the terms of the contract will be designed to minimize shirking.}

The Pauly formation implies a more general response. To the extent that the rate regulation refers to $P_H$ (the price of hospital care), not $P_M$ (the price of companion services provided by physicians), the hospital can ‘unbundle’ certain services by billing them through physicians. So long as $P_T = P_H + P_M$ is unchanged, satisfying the rate review constraint would leave quantity, quality (in the medical sense), full price, and physician net income unchanged [Pauly (1980, pp. 111–113)].

The potential adjustments to simple revenue per unit of output constraints illustrates the dilemma facing regulators. The current trend in state rate setting programs is toward greater disaggregation of case mix adjustment coupled with more aggregation of the unit of payment, e.g., use of total budget or admission bases. These more sophisticated rate review constraints reduce the hospital’s incentive to increase length of stay and to adjust service mix. On the other hand, they tend to increase the incentive for discharge-readmission stratagems and for redefining cases toward more profitable case classifications. Essentially, the regulator must balance the marginal costs of making more fine-tuned case mix determination with the welfare losses from too broad of a characterization of case mix.\footnote{We are indebted to a referee for this point.}

\section{2.5. Extensions of the model}

Recent insights by Barzel (1982) offer a means of systematically examining the changes in the hospital service bundle likely to result from a rate ceiling of the nature described here. Barzel again begins by arguing that goods are composed of measured and unmeasured attributes. Whether an attribute is measured or not depends critically on the cost of measurement. Further, once a ‘good’ is defined and transacted in the market, buyers and sellers will devote resources to exploiting the unmeasured attributes until the marginal gain is dissipated.\footnote{For example, consider a buyer and seller of oranges. Once a price per unit of a given grade has been established, the buyer has an incentive to search for those lots, within the grades, having the highest value. The seller has an incentive to provide the lowest quality lots within a given grade.}
Such activity by definition is wasteful. The firm exists, by Barzel's theory, to coordinate the incentives of the buyers and sellers, i.e., the employees, to minimize their wasteful activities in dealing with each other. Integration of the buying and selling function reduces the waste at the cost of coordination. It follows that the boundaries of a firm, i.e., the extent of integration, will be defined by the ease of measuring the relevant attributes of goods. As the costs of measuring attributes rises, the transaction will increasingly take place within the firm.

Our model suggests that the 'quality degradation hypothesis' can be made more precise. Holding other things constant, the introduction of a rate review constraint will result in the unbundling of hospital services whose contribution to the hospital product are measurable at relatively low cost — perhaps outpatient departments, pathology labs and radiology units.

A second extension of our model also follows Barzel's analysis. The coordination function implies that the firm will utilize employment contracts designed to minimize shirking activities along those dimensions that are costly to monitor. The imposition of a constraint on revenues can be expected to provide incentives to change the form of the employment contract. For example, one means of minimizing shirking in the pathology department is to tie the pathologist's remuneration to department revenues. The imposition of a rate ceiling leads to a search for means of unbundling unregulated attributes. One method, short of severing the employer-employee relationship, is to renegotiate the pathologist's employment contract to a straight salary, thereby trading the costs of pathologist output shirking for larger costs of exceeding the revenue limit. This extension yields the prediction that a rate review constraint will result in a greater incidence of salary relative to revenue (or output) based employment contracts among hospital employed physicians.

A more far-reaching extension of the general model incorporates work undertaken in organizational theory. Within that discipline, Thompson (1967) views the firm as striving to protect its 'technological core'. Economists have dismissed this allegation as inconsistent with profit maximization or, at best, reflective of a rather peculiar firm objective function. In fact, Thompson's analysis can be interpreted as cost minimization in the face of uncertainty. Indeed, major elements of Thompson's analysis are suggested in the work of Stigler (1939) and Alchian (1959).

If a rate review program is implemented with some uncertainty as to its actual constraints and duration, it is no longer obvious that a hospital would immediately alter its mix of services. One reason is the ability to cover variable but not fixed costs in the short run. A second reason, emphasized by Thompson, is, in effect, the costs of misinterpreting the actual constraint. If so, the entrepreneur optimally employs some inputs specialized in dealing
with uncertainty, or to use Thompson’s terminology, the firm ‘buffers’ the technology core. Thompson, building on Parsons (1956), views the firm as having three general functions: institutional, managerial and technical. The first is responsible for linking the firm with other organizations, including regulators. The managerial function is responsible for efficiently coordinating firm production. The technical function is the actual production of the firm’s product(s).

This notion of within-firm specialization, together with buffering the technical function in the face of uncertainty, suggests a hospital response strategy to a new rate review program. First, individual hospitals respond at the institutional level through, perhaps, direct negotiation with the regulatory agency or through appeals and lawsuits. This is followed by managerial changes, such as the addition of administrative support staff, and only then, changes in the actual hospital product [Cook et al. (1983)]. Given economies of scale in ‘buffering activities’, the increased demand for such services as a result of rate review also implies that hospitals will increasingly share such activities and, if measurement costs related to this function are large, merge their institutions.

The organizational theory rests upon the costs of adjusting to exogenous market disturbances. These adjustment costs take the form of the direct costs of altering physical plant and the mix of inputs as well as the opportunity costs of responding to inaccurately perceived market forces. Alchian (1959) maintains that unit costs can be lowered, albeit at a decreasing rate, by increasing the total volume of a production run. This result is obtained because inputs increasingly specialized to the performance of a particular task can be made to pay. One can interpret this investment in highly specialized productive capacity as Thompson’s technological core. Frequent changes in the product or the production function of a firm imply that the cost savings from production specialization cannot be achieved.

In such an environment a layered, sequential response to rate review works to smooth the firm’s rate of production, ease the rate of transition to new technology and allow collection of sequential information on the true signals embedded in the regulatory constraints. Stigler (1939) argued that changes in a firm’s production process under conditions of uncertainty depended on the duration of expected non-optimal production, the expected prices of output and the cost of adjustment. He maintained that in an uncertain environment, the firm has an incentive to forego best-known technology in favor of production processes that are passably suboptimal over a wider range of output. Finally, Stigler argued that production changes are likely to be discontinuous; he noted:

“This is not to say that alterations will necessarily be made in small increments. The minimum sensible of the entrepreneur is likely to be rather large. If output is falling, he will be loath to alter plant
accordingly, hoping for a return to 'normal' conditions — or for favorable legislation!... A critical factor is the frequency with which anticipations are revised.\textsuperscript{10}

If one expands Stigler's model to allow for inputs specialized in dealing with uncertainty one gets implications consistent with those of the organizational theorists, albeit without distinctions of institutional, managerial and technological levels of the firm, that is, a sequential response to an uncertain rate review constraint that involves increased use of inputs specialized in dealing with uncertainty and contractual changes that do not alter the hospital's production process. Only as the uncertainty is reduced does the production process change.

In summary, the ability of hospitals and physicians to alter the hospital bundle depends on (1) the operational definition of hospital services employed by the rate-setting authority; (2) the extent of insurance coverage for services received while in the hospital, but provided under independent billing; and (3) the administrative cost likely to be incurred in undertaking such changes. In general, providers will act to minimize the sum of the financial loss imposed by the price ceiling and the inefficiency loss resulting from the use of non-cost minimizing attributes and administrative structures.

This simple model seems to offer insight into the real, and much more complex, world of actual hospital rate-setting. First, it gives testable meaning to the contention that in earlier years rate-setters were 'learning by doing'. The model predicts that rate-setting authorities will add constraints through time to limit the provider's ability to inexpensively satisfy the rate ceiling in what they consider inappropriate ways. Thus, it is not surprising that New York added a length-of-stay condition to its per diem program in 1977. Nor is it unexpected that Maryland moved from per diem to per admission rate-setting provisions in 1976 and, more recently, attempted to adopt provisions dealing with hospital-based physician compensation.

Second, to the extent that providers are constrained from adopting the administratively inexpensive means of meeting the rate ceilings, services will be eliminated or added more slowly. The hospital's medical staff specialty distribution will narrow, as will its case mix. This implies reduced demands for capital, capital projects and labor. To the extent that services exhibit economies of scale, hospitals will increasingly specialize. These changes all imply a reduction in aggregate expenditures on hospital-based services.

Third, services otherwise provided in hospitals (or through hospital billings) will be provided by new organizationally freestanding providers, by physicians and by patient self-care. This new organization structure will

\textsuperscript{10}Stigler (1953, pp. 320-321). Even the reference to regulation is consistent with the Thompson view. Allowing the firm to be a demander of regulation — or in our case to be a demander of regulatory reform — is 'boundary spanning' by the 'institutional level' of the firm.
increase aggregate expenditures for non-hospital health care. The direction of the net effect on health care costs is, therefore, an empirical question.

3. The literature as a preliminary test of the model

A number of empirical analyses have addressed the issue of hospital rate setting. While none were designed to test the model presented here, they do offer an opportunity to pretest the general thrust of the theory. This section examines the effects of rate review on hospital costs, revenues, utilization, and services and contracts. It also explores the contention that different programs will have different effects. It focuses on those studies that control for confounding factors and examines several programs. The other studies have been carefully reviewed elsewhere; see Hellinger (1978), Salkever (1979), Steinwald and Sloan (1981), and Morrisey et al. (1982).

3.1. Hospital cost effects

The theory presented here argues that an effective rate review program will serve as a price ceiling and that hospitals will adjust volume and quality to equate marginal cost with the now lower revenue per unit. As a result, hospital costs per day, per admission, and per capita should increase more slowly. Sloan and Steinwald (1980) and Misek and Reynolds (1982) focused on the early 1970s and found no effects. The six major studies incorporating more recent experience found some statistically significant effects.\textsuperscript{11} However, the range of estimated effects is uncomfortably large; rates of increase in per diem costs were estimated to have been reduced by 1.6 to 6.9 percentage points, per admission rates by 1.7 to 4.9 percentage points, and per capita costs by 2.0 to 4.1 percentage points.

The range of results reflects differences in the time periods studied, the unit of analysis, the other factors considered, and the operational definition of rate review. Even in the most comparable of studies, differences in effects have alternative explanations. Sloan (1981), using 1963–78 state level data and defining rate review programs as mandatory-young and mandatory-mature, found a 3.4 percentage point reduction, short run, in the rate of increase in hospital costs per admission resulting from mature programs. Sloan (1983) extends the analysis through 1980 but uses a six-way definition of rate review. Unfortunately, one cannot determine whether the somewhat larger estimate reflects the added years of experience or the different residual categories that serve as reference groups. Less similar studies offer less chances to narrow the range of results. Coelen and Sullivan (1981), for example, use hospital specific data over the 1970–77 period, separate state

program specific measures of rate review, and a stepwise approach. Morrisey et al. (1983) use SMSA specific data, 1968–81, with separate mandatory state specific program measures and an ordinary least squares (OLS) estimation method. Coelen and Sullivan found effects per admission in Maryland, Massachusetts, New Jersey, and New York, as well as other states. Morrisey and his colleagues found effects only in New Jersey and New York. However, while these latter studies differ, they stand apart from other research because they imply major differences among mandatory programs.

The number and variety of economic studies lead to the conclusion that rate review has reduced the rate of increase in hospital costs. However, there is also evidence that the aggregate effect, as usually measured, may be driven by only a few programs.

3.2. Hospital revenue effects

The theory argues that the imposition of a rate review constraint would result in a reduction in demand as well as costs for the hospital; as a result, hospital revenues should be reduced. Three studies bear on this issue. Sloan (1981) found positive but significant mature rate review effects on the rate of increase in the ratio of total revenue to total expense over the 1963–78 period. Using 1963–80 data, Sloan (1983) found negative but insignificant mature program effects. Given Sloan’s companion findings on the reduced rate of increase in costs, the ‘profit’ finding implies that increases in revenues also were reduced as a result of rate review. Morrisey et al. (1983) examined the rate of increase in revenues per capita using SMSA data for the years 1968–81. In the five mandatory programs they examined, increases in revenues were reduced by 2.3 percent. This result was not statistically different from their cost per capita finding. However, at the individual program level only New Jersey had significant revenue effects while both New Jersey and New York displayed cost effects. Again, the thrust of our theory is consistent with the available data, but individual program differences seem important.

3.3. Utilization effects

Our model of rate review yields ambiguous results with respect to changes in hospital quantity. This result turns on the location of the preconstraint equilibrium. The widespread use of hospital insurance led us to expect that quantity would increase as a result of rate review.

Five studies have examined levels or changes in various hospital output measures as a result of rate review. Using state data Sloan (1981,1983) found no statistically significant effects on changes in either admissions or patient days. Sloan (1983) also found no effects on changes in outpatient visits or
average length of stay. However, using a shorter time period (1975–79) Melnick et al. (1981) found a 0.89 percentage point reduction in the rate of decrease in state specific average length of stay. Their total admission results corroborate Sloan’s. Joskow (1980) examined a 1976 cross-section of hospitals to determine the effects of rate review on unoccupied beds per day. Using OLS, he found a 20 percent reduction in this variable, i.e., greater occupancy. However, his weighted least squares analysis yielded no statistically significant effects. Unfortunately, for our analysis, Joskow does not control for the number of beds.

The most extensive examination of utilization to date was conducted by Worthington and Piro (1982). Using the analytic design of Coelen and Sullivan (1981) but in an OLS framework, they found that rate review generated statistically significant increases in occupancy rates in 11 out of the 25 programs examined. There were no statistically significant reductions. Average length of stay was estimated to have increased in eight programs but decreased in one. Admissions per bed were estimated to have increased under four programs, but decreased slightly in one. However, this study failed to control for the number of beds. Thus, while no study refutes our theory and empirical presumption, the data supporting it are not ideally suited to the test.

3.4. Service and contract effects

We have argued through the model and its extensions that effective rate review programs would lead to reductions in hospital services, the spinning-off or unbundling of services, and changes in contractual arrangements. The literature offers five studies that support these issues.

Joskow (1981) cross-sectionally examined the diffusion of computer augmented tomography (CT scanners). Controlling for certificate of need legislation and a few other factors, he found no effects of rate review on the number of CT scanners in the state. However, rate review did result in a small but statistically significant increase in the proportion of CT scanners in physician’s offices. To our knowledge, this is the sole empirical analysis of unbundling.

Steinwald (1983) examined the form of compensation for hospital-based physicians in 1979. Using a logit technique, he examined the probability of ‘percent of departmental revenues’ and ‘salaries’ relative to ‘fee-for-service’ compensation for anesthesiologists, pathologists, radiologists, cardiologists and emergency medicine specialists. Using a dummy for states with mandatory rate setting programs he found statistically significant effects for four of the five specialties. In general, percentage of revenue declines but salary increases relative to fee-for-service. These results are broadly consistent with our theory but are not direct tests because Steinwald combined
physicians billing directly with those billing through the hospital for each form of compensation.

Cromwell and Kanak (1982), using the Coelen and Sullivan (1981) methodology, examined the effects of specific rate setting programs on the services and facilities offered by hospitals. In their first approach, services were grouped into categories such as quality enhancing, complex, and community and were used in a standard diffusion model, with a lag term, to examine the number of services offered in each group. While finding a few statistically significant reductions (and increases) in service adoption as a result of rate review, the approach generally found no effects. Their second approach used a linear probability model, without a lag variable, to examine 13 selected services. In the category of complex services, the 25 programs yielded several statistically significant effects at the 90 percent level: radiotherapy, two reductions, one increase; intensive care units, seven reductions, one increase; open-heart units, four decreases; EEG capability, five decreases, one increase; radio-isotop capability, two reductions; burn-care units, three reductions, three increases; and physical therapy, three programs had statistically significant reductions in the probability of a hospital having the capability. Examined from another perspective, the New York program generated statistically significant reductions for six of these seven services; New Jersey was estimated to have reduced four while increasing one. The effects in other programs were less pronounced.

In a companion study, Kidder and Sullivan (1982) examined hospital full-time equivalents (FTEs) per patient day, among other payroll and labor issues. While changes in this variable may be offset by changes in capital per day with no net change in services, FTEs per patient day can be a reasonable proxy for services provided if capital stock is largely fixed. Kidder and Sullivan found that seven of their 25 rate setting programs yielded negative and statistically significant reductions. The coefficients ranged from −0.03 in Massachusetts to −0.10 in Maryland.

These results, while not overpowering, are consistent with our theory. They provide evidence of unbundling, retracting and service reductions. A recent study by Morrissey et al. (1983) suggests that it may not be a trivial matter to effectively move services from hospitals' to physicians' bills. That study examined the effects of rate review programs on Medicare expenditures, parts A and B separately. The analysis found smaller, but statistically significant, reductions in part B (physician) expenditures than in part A (hospital) expenses per enrollee. These findings may not be generalized as far as the larger population, but they do suggest that the production process and the ease with which physicians can bill in the

12The authors use three alternative grouping schemes; they place most emphasis on the scheme developed by Berry (1973).
hospital setting may make unbundling profitable along only a relatively small number of margins.

3.5. Different programs—different effects

Our theory has argued that different programs will have different effects on hospital behavior, because different rate review programs constrain different margins of adjustment. The preceding discussion of the findings of the literature has glossed over the differences among programs, because the researchers have tended to aggregate programs. Such aggregation can produce misleading results.

Morrisey et al. (1983) is the only study to have explicitly considered alternative definitions of rate review. Their cost effects using a single variable for mature mandatory programs mirrors the findings of others; however, at a disaggregated level the results are driven by New Jersey and New York. The works of Coelen and Sullivan (1981), Worthington and Piro (1982), Cromwell and Kanak (1982), and Kidder and Sullivan (1982) attest to the rather large differences in effect that result from the 25 programs they examine.

Taken as a whole, the existing empirical literature offers only two clear insights into the type of programs that have been effective in controlling costs. First, the results suggest that no program has been successful in reducing hospital costs in the first two or three years of operation. This phenomenon may be the result of learning by doing or of studied responses by hospitals and physicians as we have suggested. Second, most programs have had smaller cost effects than have mandatory state run programs [Sloan (1983), and Coelen and Sullivan (1981)] and typically smaller non-cost effects [Worthington and Piro (1982), Kidder and Sullivan (1982), and Cromwell and Kanak (1982)]. These conclusions may reflect non-binding programs or a variety of low-cost adjustment opportunities. More generally, however, the literature is broadly consistent with our model. The studies taken together suggest differential but widespread effects on the operations and organization of the hospital as a result of rate review programs.

4. A research agenda

The preliminary tests from the literature indicate how far researchers have to go to understand the effects of rate review programs. This section discusses four areas for further research.

4.1. Regulation and its measurement

It is commonly alleged that rate review programs cannot be transported to different political environments and expected to have similar results. This
argument has two components, each worthy of research. First is the endogeneity issue. Does a rate review program have an independent effect or is it purely a function of the underlying regulatory preferences and socioeconomic characteristics of the community? Edward’s (1978) examination of compulsory education suggests the direction of such an investigation might take. Second, how dependent are rate review effects on the other constraints imposed on the industry? If the effects depend on the margins of unconstrained behavior, then the effects depend on the limits imposed by state licensure, planning bodies, and utilization review programs. Transporting rate review alone would be insufficient.

A second more technical area for further research is the measurement of regulatory programs themselves. First, some aggregate measure of ‘regulatory intensity’ would be useful. Description studies [Hamilton et al. (1980)] indicate the multidimensional nature of rate review programs, yet empirical researchers have measured only one or two dimensions and have missed potentially important variations. Second, our theory suggests further research on the identification of theoretically meaningful criteria for ranking programs.

We have argued elsewhere [Cook et al. (1983)] that the intensity of a particular regulatory program is determined by a set of underlying characteristics of that program: (1) the scope of the institutions and the dimensions of institutional behavior constrained by the regulation; (2) the stringency of the program in terms of the restrictiveness of the rate review constraint and the level of its enforcement; (3) the maturity of the program, which reflects the extent of learning by doing by the regulatory body and which indicates the particular stage of the life cycle of response to regulation which hospitals are likely to occupy at any given point in time; and (4) the degree of uncertainty regarding the content and enforcement of the rate review program.

Urban and Bice (1981) have developed and tested a similar approach. In their model, regulatory intensity is measured as the joint product of such factors as the program’s scope (number of payers and hospitals covered), the use of the appeals process for rate review decisions, the unit of payment, and locus of program control. They also identify some of the exogenous determinates of regulatory intensity — a particularly useful outcome for researchers wishing to develop instrumental variable estimators of regulatory intensity. Their results suggested that the principal determinates of rate review intensity were the ‘innovativeness’ of the state (based on a measure of legislative activity), the number of hospitals in the state, Medicaid expenditures, and per diem hospital costs prior to the period of observation. Clearly, more work remains to be done before researchers can structurally identify the underlying determinates of regulatory intensity.

In addition, future research should attempt to partial out the independent
and interactive effects of scope, stringency, maturity, and uncertainty of programs. For example, our theory implies that a program that constrains revenue per patient day creates an incentive for increased length of stay, whereas a program that specifies revenue per admission will induce decreased length of stay and a shift toward less complex cases, ceteris paribus. Such differential predictions related to program scope can only be tested if the stringency of the program — as measured, perhaps, by the availability of retrospective adjustments — and the level of resources devoted to enforcement are controlled simultaneously.

It is particularly important that future studies include measures of the uncertainty and instability of the regulatory program. Not only does uncertainty concerning the content and enforcement of the program mitigate hospital response to the underlying regulatory constraints, it also increases the potential for hospitals to negotiate with the agency for substantive changes in the program itself. None of these factors have been addressed explicitly in previous empirical work.

The extent of operating coordination among particular regulatory programs, especially the state rate review and certificate of need (CON) programs, is a critical element in the measurement of regulation. While earlier work has included separate rate review and CON variables, our analysis suggests interaction effects, for example, as CON constrains the service mix adjustments hospitals would otherwise make in response to rate review. These interactions should be tested explicitly.

Enforcement and the range of available sanctions are key determinates of the stringency of rate review programs. Preliminary research [Chapko (1983)] reveals several operational indicators that load on an hypothesized common factor of enforcement: requirements that the hospital repay any excess revenues, the proportion of excess revenues that must be repaid, the number of years over which the hospital may spread repayment, and financial penalties for failing to attain minimum utilization (occupancy standards and/or exceeding maximum length of stay standards). Measuring the actual extent of enforcement of the rate review program is essential because incentives to avoid the spirit and perhaps the letter of the regulatory program increase as the constraint is tightened.

Finally, our analysis has relied on traditional comparative statistics; future work should consider dynamic adjustment implications. If regulators adjust to hospital behavior with a lag, e.g., by setting rates on the basis of allowable costs in the base year, they may produce cycles in hospital revenue, costs and output. In a general analysis of public utility regulation Mann (1979) has illustrated that — for the plausible case of decreasing returns to scale and some price elasticity of demand — regulatory lag will generate a cyclical trend toward increasing costs, rates and output initially, but the combination of decreasing returns and price sensitivity will ultimately produce a
dampening effect on costs, rates and output. On the other hand, his cycles vanish under conditions of constant marginal costs.

4.2. Hospital operations and organization

The available evidence, and certainly the theory, suggest that the effects of rate review on hospitals can be pervasive. Research should be directed to identifying the effects on utilization, service mix, patient and physician mix, contracting behavior, and service sharing. Further, analysis must be directed to the effects of rate review, and regulation more generally, on the formation of multi-institutional systems, and on the administrative and medical staff organizations of the hospital. Further, explicit empirical tests of the time path of individual hospital response to regulation would be extremely useful in understanding how and why rate review affects costs.

Second, there has been no systematic empirical study of the differential effect of rate review programs on non-profit vs. for-profit hospitals. Careful examination of differences in the time path of response to regulation and the magnitude of those responses should be conducted. For example, one might predict that service mix changes induced by regulation would tend toward lower cost services in both types of hospitals, but that the negative 'income effect' of regulation on physicians-qua-residual claimants in non-profit hospitals, in addition to the substitution effects favoring less costly services, would induce a stronger shift among the non-profits toward physician income preserving services relative to lower cost services. In the for-profit hospital only the substitution effect toward lower cost (higher profit) services would be observed.

Third, further research should be focused on the effects of rate review on hospital internal organization and, in turn, the effects of internal organization on hospital costs. Sloan and Becker (1981), while not addressing rate review do demonstrate the potential importance of internal organization in explaining the behavior of hospitals. They found, for example, that expanding clinical departmental responsibilities without implementing cost minimizing incentives will probably raise costs. More generally, several of their organizational variables have statistically significantly effects on hospital costs per admission and per day. Similar studies applied to rate review are essential if one is to learn how and if hospitals become more efficient as a result of rate review programs.

Finally, recent developments in the theory of the multiproduct firm [Baumol et al. (1982)] suggest alternative approaches to the analysis of the effects of rate review on the operations and costs of multiproduct hospitals. Our analysis focuses on the vector of attributes and characteristics comprising the hospital bundle of services. This approach should be generalized into the vectors comprising the respective hospital products. This
generalization has the advantage of focusing on the economics of scope, i.e., the complementarity in joint production which is only implicit in our model. Such complementarity is presumably at the heart of our unbundling thesis.

4.3. Non-hospital effects

The unbundling of the hospital product as a result of rate review suggests that important effects can be found outside the hospital sector. Physician office visits and hospital outpatient visits may increase. The prevalence of major capital equipment in medical offices or freestanding firms should increase as a result of rate review programs directed at hospitals. Similarly, hospital corporate restructuring would be expected to increase. The presence of these effects has not been determined, nor have any cost savings that may result from them.

The current interest in multi-institutional arrangements poses several empirical questions for the study of the effects of rate review on the hospital market. Specifically, if hospital systems have a comparative advantage in buffering the institution from regulatory uncertainty, then systems will respond more slowly than independents at the technical core level. In addition, one might posit that the current trend toward hospital diversification into ambulatory surgical centers, home health care and other ‘non-hospital’ functions will be reinforced by increasing regulatory intensity. Detailed statistical analysis of the effect of rate review on these phenomena is sorely needed.

4.4. Overall cost effects

If the hospital product is being unbundled as a result of rate review, cost savings measured solely in the hospital sector will overstate the true impact of rate review programs because some functions will now be performed outside the hospital. Others will simply be removed from the hospital budget. These services entail costs that mitigate some or all of the hospital savings. On the other hand, complementarity between hospital and physician services implies that effective hospital rate review programs have reduced the rate of increase in physician revenue as well. The relevant question for policymakers is the overall impact of the program on health care costs. One approach would examine total episode costs of care for specific diagnoses under alternative regulatory settings. A second, and much more complicated analysis, should be directed at an aggregate benefit–cost analysis designed to determine the net effect of rate review programs on health care costs. Only with this sort of aggregate investigation will it be possible to arrive at an overall evaluation of rate review programs.
References


Parsons, T.E., 1956, Suggestions for a sociological approach to a theory of organizations, Administrative Science Quarterly 1, 63–85.


Salkever, David, 1979, Hospital sector inflation (Lexington Books, Heath, Lexington, MA).


Stigler, George, 1939, Production and distribution in the short run, Journal of Political Economy 47, no. 3, June, 305–327.

