Improving Economic Incentives in Hospital Prospective Payment Systems Through Equilibrium Pricing

Michael Shwartz • Melanie L. Lenard
School of Management, Boston University, 621 *Commonwealth Avenue, Boston, Massachusetts* 02215
School of Management, Boston University, 704 *Commonwealth Avenue, Boston, Massachusetts* 02215

Under the Prospective Payment System (PPS) implemented by Medicare in 1983, hospitals are paid a set price for each Medicare patient treated, rather than being reimbursed for the patient's costs as had been done previously. An increasing number of other insurers have adopted a similar method of hospital payment. In these systems, the price, which depends on the patient's Diagnosis Related Group (DRG), is derived from the average cost over all hospitals of all patients in that DRG. We propose an alternative method for setting prices in hospital prospective payment systems, called equilibrium pricing, in which prices are derived from a linear programming model of competitive equilibrium. To evaluate the improvement in incentives associated with equilibrium pricing, we define a measure, called the disincentive index, of the extent to which a set of prices creates economic disincentives to efficient behavior. In the situation in which all hospitals compete in a single market area, we show that equilibrium pricing creates the best possible economic incentives, i.e., by reducing the disincentive index to zero. The analysis is then extended to the more realistic situation where hospitals compete in limited geographical market areas, whereas prices must be uniformly set for a number of such market areas. We prove that, with an appropriate generalization of the disincentive index, equilibrium prices for a single market area are also optimal for multiple market areas. Finally, actual cost and utilization data from hospitals that compete in eastern Massachusetts are used to determine prices and to evaluate the associated disincentive index for a simulated prospective payment system. This empirical study shows a dramatic improvement in the incentives created by equilibrium pricing compared to average-cost pricing.
(*Hospital Payment Systems; Economic Equilibrium Models*)

Medicare is the federally financed health insurance program for the elderly. Prior to October 1983, Medicare reimbursed hospitals on the basis of their costs. The rapid rise in resources devoted to health care following enactment of Medicare in the mid 1960s has been attributed in part to this method of cost-based reimbursement, a method which created little incentive for efficiency.

In October 1983 Medicare changed its method of reimbursement when it implemented the Prospective Payment System (PPS). Under the PPS, hospitals are paid a price for each patient, rather than being reimbursed for the patient's costs. The price established for each patient is derived from the average cost of treating patients with "similar" diagnoses. An increasing number of state Medicaid programs and private insurers have adopted a similar method of prospective payment.

No justification has ever been provided for setting prices for different types of patients based on the average cost of treatment. Further, there has been little

0025-1909/94/4006/0774$01.25

discussion of methods or measures for evaluating alternative pricing policies (e.g., using median costs as opposed to average costs). In this paper, we propose an approach for both establishing and evaluating prices that is based on the extent to which the prices motivate hospitals to act in an economically efficient manner.

In what follows, we first provide background on the PPS and then provide a conceptual overview of our approach. In §3, we consider the problem of establishing prices for the case where all hospitals compete in a single market area. We interpret the classic linear programming model of economic equilibrium (Gale 1960) to fit this situation, and show that if prices are derived from the dual prices associated with this linear programming model, an approach we will refer to as equilibrium pricing, incentives are created for economically efficient behavior. In §4 we propose a measure, called the disincentive index, of the extent to which an arbitrary set of prices creates economic disincentives to economically efficient behavior. In §5, we extend the analysis to the situation in which groups of hospitals compete in different market areas. In §6, actual cost and utilization data from 63 hospitals that compete in eastern Massachusetts are used to determine equilibrium prices and to compare the incentives created by equilibrium prices to those under average-cost prices. In §7, we discuss our overall approach, the models and our results.

1. Background on the PPS

The basis for payment in the PPS is the patient's Diagnosis Related Group (DRG). DRGs, created by a research group at Yale University (Fetter et al. 1980, Fetter 1991), were an attempt to use a combination of statistical analysis and clinical judgment to group the over 10,000 ICD-9-CM diagnostic codes (International Classification of Diseases, Ninth Revision, Clinical Modification) so that clinical coherence was preserved and resource use was relatively homogeneous within each DRG. DRGs (currently there are 490) are distinguished based on whether or not a surgical procedure is performed, principal diagnosis, complications and comorbidity, discharge status, and (until recently) age. Under the PPS, a hospital is reimbursed a price derived from the average cost of treating patients in each DRG, regardless of what the patient's actual costs are.

The change from cost-based reimbursement to prospective payment was part of an effort during the Reagan Administration to control health care costs by introducing the "rigors" of the marketplace. As stated by Richard Schweiker, Secretary of Health and Human Services, in a report to Congress in December 1982, "From the hospital's point of view, prospective rates represent a set of prices with similar characteristics to the prices it would face in a more conventional market. The hospital knows the amount it will be paid per unit of service and that the payment rate will remain unchanged regardless of its own cost experience. Thus, like firms in other markets, the hospital bears the risk that the prospective payment rate will not cover its cost per unit of care. In general, this risk generates strong financial incentives for hospitals to control resource use" (Schweiker 1982, p. 15).

Later in the report, he states "After determining what allocation of resources within the hospital's existing framework and case mix will achieve the lowest cost for each DRG . . . , the hospital will deliberate on output decisions for each patient type to make its overall case mix volume as productive as possible. Thus, hospitals will tend to specialize more under PPS than they currently do" (Schweiker 1982, p. 103).

In the PPS, the price for each DRG was originally derived from a set of relative weights that were calculated from the average cost of Medicare patients in the DRG compared to the average cost of all Medicare patients. Currently, prices are derived from weights calculated from average charges, based on a study (Cotterill et al. 1986) showing that cost-based weights and charge-based weights are similar (though recently this relationship has been challenged (Price 1989)). Prices for each DRG are established by specifying a price per unit of relative weight. Since the distribution of patients across DRGs in any year can be estimated from data from the previous year, this method allows prices to be established and modified in a way to constrain the overall size of the Medicare budget. Originally, prices were set to achieve budget neutrality, that is, so that estimated total expenditure would be the same under the PPS as it had been under cost-based reimbursement.

The relative weights are recalculated yearly and new prices established. In this way, DRG prices reflect changes due to new technology, improved hospital efficiency, favorable returns to scale that may be associated with increased specialization, and any increases in system efficiency if more patients are treated at more efficient hospitals.

A variety of State Medicaid programs and private insurers that have adopted forms of DRG-based per case prospective payment use an approach similar to the PPS to set prices. In particular, in all of these systems, prices are based on average costs or average charges, something we will refer to as average-cost pricing.

2. Conceptual Overview

As noted by Schweiker, there are two types of economic incentives created by a prospective payment system: one, to control resource use in general; and two, to specialize in those DRGs where the hospital is relatively most efficient. Any set of DRG prices will create incentives to reduce costs. However, different sets of prices will differ in the extent to which certain DRGs are more or less economically attractive and thus in the amount and type of specialization encouraged.

We propose an approach for both establishing and evaluating prices that is derived from explicit acknowledgment of the type of specialization behavior society would like to encourage. Specifically, if the "best" or optimal allocation of patients to hospitals were known, prices should be determined so that economic incentives are created for hospitals to act in a way consistent with this allocation; alternative sets of prices should be compared in terms of the extent to which they motivate behavior consistent with this allocation.

Determining the optimal allocation of patients to hospitals is an extremely complex undertaking. A wide range of factors should be considered, including the efficiency with which different hospitals treat different types of patients, the quality of care provided to different types of patients, relationships between volume and both cost and quality, geographical access of patients to hospitals, and patient and physician preferences. Since pricing policies attempt to influence the behavior of hospitals through economic incentives, our approach to determining the optimal allocation of patients to hospitals emphasizes economic criteria. Other criteria, such as meeting demand for treatment, maintaining quality of care and geographical access, are incorporated by means of constraints.

In what follows, we formulate two linear programming models. The first is a model of the problem from society's perspective; it determines the allocation of patients to hospitals which will minimize costs subject to certain constraints. The second is a model of the problem from an individual hospital's perspective; it determines for a given set of prices what mix of patients they would choose in order to maximize their profits. The allocation determined by society's model, which does not depend on any pricing structure, will be called the "economically efficient allocation"; behavior consistent with this allocation we will call "economically efficient behavior." The economically efficient allocation is also a competitive equilibrium allocation, i.e., the allocation that would result if hospitals operated in a theoretical world of "pure" competition.

Solution to the individual hospitals' models will differ for different sets of prices. To the extent that a set of prices motivates hospitals in the aggregate to exhibit economically efficient behavior, the prices are "good"; to the extent it does not, prices are "bad."

The problem with "bad" prices is that they create unequal economic incentives for treating different types of patients. Patients in certain groups of DRGs are overly profitable, motivating hospitals to provide too much capacity to treat these patients relative to the number available. For other patient-types, the economic incentives to treat them are too low, motivating hospitals to reduce capacity for these patients. The result is that, though hospitals may be encouraged to specialize by "bad" prices, they will not be encouraged to specialize in a way consistent with overall system efficiency.

Currently, most hospitals have excess capacity. Since the hospital industry is characterized by high fixed costs, any patient for whom revenue is expected to exceed short-term variable costs will be desirable. Thus, patients in less economically attractive DRGs are unlikely to find access limited. The greater problem is rather with the overly profitable DRGs. Since expansion of capacity often involves expensive technology or construction, the costs associated with overexpansion are likely to burden the health care system for a long time. In addition, unused capacity may create incentives for supply-induced demand (Wennberg 1982). Thus, incentives to overexpand certain types of capacity are likely to interfere with the cost containment goals of the PPS.

In the next section, we describe the details of the linear programming model we propose for determining the optimal patient allocation and DRG prices. As will become apparent, our focus in this model is not on all hospitalized patients, but only on some percentage of patients whose hospital-using behavior is subject to economic incentives, incentives that may influence patients directly or have an effect through their physician. Thus, when we talk about a reallocation of patients, we are only talking about some perhaps small percentage of total patients; when we talk about prices influencing behavior, we are talking about an influence that occurs only at the margin.

Though ultimately a price will be established for each DRG, hospitals cannot usually control the number of patients treated in individual DRGs. Rather, management decisions are usually made about the size of clinical subspecialties, or about whether or not to offer specific treatment modalities within clinical subspecialties. Thus, in evaluating economic incentives created by alternative pricing systems, it is important to focus on aggregations of DRGs that are relevant for management decision-making.

We refer to the groups of patients resulting from the aggregation of DRGs into units about which management decisions might reasonably be made as patient-types. Since our concern is that correct incentives exist in regard to patient-types, the number of patients of each patient-type treated at each hospital is the decision variable in the linear programming model. The output of the analysis is a price for each patient-type. DRG prices would be calculated based on relative cost within patient-type, as is done in the current Medicare PPS (though in the PPS relative cost weights are calculated over all patients, whereas in our case relative cost weights would be calculated from the average cost of patients in the DRG relative to the average cost of patients in the patient-type group).

3. The Single Market Area Problem

In this section, we formulate a linear programming model of the problem of reallocating patients in order to minimize overall costs of treatment. We take this as society's view of the problem. We show that if prices are derived from the dual variables associated with the societal model, economically efficient incentives are created.

Assume N hospitals compete for patients in one market area, that is, patients in the area can travel to and be treated at any of the hospitals in the area. Assume further that each patient belongs to one of M patient-types. For each patient-type and for each hospital, the average cost of treatment is known.

Because prices are recalibrated yearly, it is appropriate to take a short-term perspective. Each hospital will be constrained to function at its current annual caseload, which is assumed to reflect its ability to compete in the marketplace at this time. Further, it is assumed only some (perhaps small) percentage of this current caseload is available for "re-allocation" to a different case-mix. This is the hospital's capacity available for reallocation. Similarly, the demand for treatment of each patient-type is assumed fixed in the short-term; only some (perhaps small) percentage of the current total caseload in the market area for each patient-type is assumed available for reallocation to different hospitals.

Let

x_{ij} = the number of patients of type j treated at hospital i;

C_{ij} = the average cost per patient for hospital i to treat patients of type j;

K'_i = the number of patients of all types currently treated by hospital i;

D'_j = the number of patients of type j currently treated by all hospitals in the area.

Now, let

α = percentage of patients available for reallocation; then,

$K_i = \alpha K'_i$ = number of patients available at hospital i for case-mix changes;

$D_j = \alpha D'_j$ = number of patients of type j available for reallocation.

Since $\sum_j D'_j = \sum_i K'_i$ = the total number of hospitalized patients in the area, the above definitions ensure that the total hospital caseload available for reallocation equals the total number of patients available for reallocation, or $\sum_j D_j = \sum_i K_i$. Furthermore, prices derived from this model will be such that each hospital can compete successfully at its current size (defined by the number of patients it currently treats) and that all patients of every patient-type will be economically attractive to some hospital in the area. Also, it should be noted that our primary results (prices and comparative measures of prices) do not depend on the value of α, though the number of patients reallocated and the economic impact of the reallocations will be proportional to α.

Taking a societal perspective, we would like a price system that encourages hospitals to focus on patient-types in a way that minimizes total treatment costs subject to patient availability and the caseload constraint at each hospital. The minimum cost allocation can be found from the solution to the following linear programming model: select values for x_{ij} to

$$\begin{aligned} \text{minimize} \quad & \sum_i \sum_j C_{ij} x_{ij} \\ \text{subject to} \quad & \sum_i x_{ij} \geq D_j, \quad \text{for each } j, \quad j = 1, \ldots, M; \\ & \sum_j x_{ij} \leq K_i, \quad \text{for each } i, \quad i = 1, \ldots, N; \\ & x_{ij} \geq 0, \quad \text{for all } i \text{ and } j. \end{aligned}$$

We refer to this model as the societal optimization model, or Model S. x_{ij}^{*s} is the optimal value of x_{ij} in Model S.

An appropriate set of prices can be found by solving the dual of Model S. Let the variables in the dual problem be r_j and s_i. The dual of Model S is

$$\begin{aligned} \text{maximize} \quad & \sum_j r_j D_j - \sum_i s_i K_i \\ \text{subject to} \quad & r_j - s_i \leq C_{ij}, \quad \text{for all } i \text{ and } j; \\ & r_j \geq 0, \quad \text{for each } j, \quad j = 1, \ldots, M; \\ & s_i \geq 0, \quad \text{for each } i, \quad i = 1, \ldots, N. \end{aligned}$$

Consider setting the prices p_j for each patient-type equal to the optimal values of the dual variables associated with the demand constraints, i.e., $p_j = r_j^{*s}$. From the complementary slackness relationship between the solution of Model S and its dual, we know that for any j where $x_{ij}^{*s} > 0$, then $r_j^{*s} - C_{ij} = s_i^{*s}$. Thus, s_i^{*s} is the profit hospital i will earn per case on any patient-type. If $r_j^{*s} - C_{ij} < s_i^{*s}$, then $x_{ij}^{*s} = 0$. Those patient-types that we would like hospital i to treat in order to minimize cost to society ($x_{ij}^{*s} > 0$) are exactly the ones it wants to treat because they are the most profitable (they all have profit s_i^{*s}). The ones it has no incentive to treat (since its profit is less than s_i^{*s}) are those we do not want hospital i to treat ($x_{ij}^{*s} = 0$).

Budget neutrality can be incorporated as follows. Since $\sum_j D_j = \sum_i K_i$, the optimal values of the dual variables r_j and s_i are not unique. If r_j^{*s} and s_i^{*s} are optimal, so are $r_j^{*s} + c$ and $s_i^{*s} + c$, where c is an arbitrary constant. If B is the total size of the health care budget, choose c so that $\sum_j (r_j^{*s} + c) D_j = B$. In particular, $c = (B - \sum_j r_j^{*s} D_j) / \sum_j D_j$. Let $p_j^{*s} = r_j^{*s} + c$. We will refer to the set of prices $\{p_j^{*s}\}$ as "equilibrium prices."

The optimal solution of Model S and its dual is in fact the competitive equilibrium (Koopmans 1951). For any set of budget-neutral prices, total revenues to the industry will be the same. Thus, the minimum cost solution to Model S is also the maximum profit solution for the industry as a whole.

We have shown that because of the complementary slackness relationship between a primal and its dual, hospitals will be indifferent among those patient-types that they treat under the optimum allocation. Over time, each hospital will discover the limit on its ability to increase the number of patients of each patient-type it would like to attract, and thus the industry will eventually settle into the equilibrium allocation (Adam Smith's "unseen hand"). However, in the short run, hospitals will not know the number of patients of each patient-type they should treat under this allocation. Thus, as Dantzig has pointed out (1963, pp. 464–465), even in a theoretical world of pure competition, there is a useful role for a benign regulator, who, by suggesting production quotas for different goods at each producing unit, could reduce short-term inefficiencies.

So far, we have mainly restated in the context of our problem well-known results about linear programming and competitive equilibrium. (For example, see Gale 1960, pp. 85–92, who states at the end of this section "For a competitive economy in which each firm operates a set of linear activities it turns out that it is possible to assign prices to resources in such a way that, although each firm acts so as to maximize its profits, the demand for resources will not exceed available supply. Furthermore, at these prices the firms in maximizing their own profit will automatically be operating so as to maximize the value of the total output of the economy.")

In what follows, we extend these results in two ways.

First, we develop a measure of the quality of the incentives created by an arbitrary set of prices. Then, in §5, we develop a multi-market area model which takes into account limits on geographical access to hospitals.

4. An Index of Economic Disincentives

In response to a set of prices, hospitals will attempt to increase their profit by reallocating a portion (assumed to be α) of their caseload. When each hospital's most profitable reallocation corresponds to society's least cost reallocation, hospitals will not only be acting in their own best interests but minimizing total treatment costs. This suggests that a measure of the quality of incentives created by a set of prices should be derived from the difference in profit potential between the optimum reallocation from society's perspective and the aggregation of the most profitable reallocations from each hospital's individual perspective.

We begin by formulating a profit-maximization model for each individual hospital. As in the case of society's model, define the decision variables x_{ij} as the number of patients of type j treated at hospital i. Further, let p_j = price paid per patient for patient-type j. Then, for fixed $i = I$, hospital I's optimization problem is to select values for x_{Ij} which

$$\begin{aligned} \text{maximize} \quad & \sum_j (p_j - C_{Ij})x_{Ij} \\ \text{subject to} \quad & \sum_j x_{Ij} = K_I; \\ & 0 \le x_{Ij} \le D_j, \quad \text{for each } j, \quad j = 1, \ldots, M. \end{aligned}$$

We refer to this model as Model h_I. The solution to Model h_I will be called a hospital's most profitable allocation and will be denoted by x_{Ij}^{*h}. Since each hospital decides individually what to do, there is nothing to ensure that, in aggregate, the number of patients of a given patient-type desired by all hospitals in the area corresponds to the number that exists in the area.

For a given set of reimbursement prices $p = \{p_j\}$, let $Z_i^{*h}(p) = \sum_j (p_j - C_{ij})x_{ij}^{*h}$ = maximum profit hospital i can achieve with its most profitable allocation;

$a_i(p) = Z_i^{*h}(p)/K_i$ = maximum profit *per case* hospital i can achieve with its most profitable allocation;

$O_{ij}(p) = a_i(p) - (p_j - C_{ij})$ = opportunity cost per case if hospital i treats patient-type j;

$OC_j(p) = \sum_i x_{ij}^{*S} O_{ij}(p)/D_j$ = average (over all hospitals) opportunity cost per case associated with patient-type j under the societal optimal allocation.

$O_{ij}(p)$ is a measure of the economic disincentive for hospital i to treat patient-type j. If $O_{ij}(p)$ is small, patient-type j is almost as attractive (or perhaps even more attractive since $O_{ij}(p)$ could be negative) to hospital i as its most profitable mix of patient-types, and thus hospital i has an economic incentive to treat these patients. If $O_{ij}(p)$ is large, hospital i has little incentive to treat patient-type j. Thus, $O_{ij}(p)$ may be thought of as the economic disincentive for hospital i to treat patient-type j.

$OC_j(p)$ is an overall measure of the disincentives to treat patient-type j. When $OC_j(p)$ is small, those hospitals that treat patient-type j under the societal optimal allocation (hospitals with a nonzero value for x_{ij}^{*S}) have little disincentive to treat these patients ($O_{ij}(p)$ is small). Under these circumstances, we say that desirable incentives are created by the prices. When $OC_j(p)$ is large, the opposite occurs, i.e., there is a lack of concordance between the societal optimal allocation and each hospital's most profitable allocation, and thus disincentives exist to treat patient-type j.

For any set of budget neutral prices $p = \{p_j\}$, we can define $V(p)$, the overall measure of economic disincentives, to be the average per case disincentive averaged over all patient-types, or

$$V(p) = \sum_j OC_j(p)D_j \Big/ \sum_j D_j.$$

Note that $V(p)$ may also be written as

$$V(p) = \sum_i \sum_j O_{ij}(p)x_{ij}^{*S} \Big/ T, \tag{3.1}$$

where $T = \sum_i K_i = \sum_j D_j$.

Thus, $V(p)$ is the average opportunity cost per case over all hospitals and all patient-types when society's optimal allocation is attained.

In Appendix A, we show that $V(p)$ is the difference between the maximum average per case profit hospitals could achieve if each hospital realized its most profitable allocation (i.e., its solution to Model h_I) and the maximum average per case profit hospitals could achieve under the societal optimal allocation (i.e., the solution to Model S). We also prove in Appendix A that equilibrium prices create the best possible incentives by reducing $V(p)$ to zero. Finally, we prove that with equilibrium prices, society's best allocation will also be the most profitable allocation for each individual hospital.

5. The Multi-market Area Formulation

So far we have only considered the case in which hospitals compete in a single market area. The single market model does not reflect the fact that most patients are unlikely to attend a hospital that is not geographically accessible. This creates a problem if the geographical area is large because hospitals that find a particular patient-type profitable may be far from where many of the patients live. There is a variety of ways in which geographical access might be incorporated in the model formulation. We prefer an approach that explicitly recognizes multiple market areas, where markets are defined so that hospitals that compete in a market area are reasonably close to patients in the area (under 30 minutes in the empirical study in §6).

Hospital market areas are usually defined based on geographical considerations and patient origin data. However, there is no generally agreed upon approach for actually defining boundaries and for assigning hospitals to specific areas (Garnick et al. 1987). Therefore, it is unreasonable to consider pricing policies that depend in an important way on how the market areas are defined, e.g., one that pays different prices depending on market area. In the analysis that follows, we use the multiple area model to determine a least cost competitive equilibrium patient allocation that takes into account the geographical proximity of hospitals to patients. This allocation is then used to evaluate incentives associated with alternative pricing methods.

In the linear programming formulation, we maintain geographical access by adding market area constraints to the linear programming model. The purpose of these constraints is to ensure that the group of hospitals that compete in a market area devotes sufficient capacity (at least as much as they had provided last year) to patients that reside in that area.

It simplifies both the formulation of the model and the interpretation of results to assign each hospital to only one market area. Since in most cases a high percentage of a hospital's caseload comes from its primary market area (in the empirical study that follows, in 56 of 63 hospitals over 80% of a hospital's caseload was from its primary area) and since the market areas are used only to approximate reasonable geographical access, little is lost by this simplification.

Let

L = number of market areas;

H_k = the set of hospitals that compete in market area k, $k = 1, \ldots, L$ $(H_k \subset \{1, \ldots, N\})$;

A_i = the market area in which hospital i competes, $A_i \in \{1, 2, \ldots, L\}$

D'_{jk} = the number of patients of type j who reside in market area k, and were treated (last year) by hospitals that compete in area k;.

$D_{jk} = \alpha D'_{jk}$ = the number of patients of type j who reside in market area k, were treated (last year) by hospitals that compete in area k, and are available for reallocation.

Let D_j and K_i be defined as in the single market area problem. It is important to note that, for every j, $D_j > \sum_k D_{jk}$ (since some patients are treated by hospitals outside of the area where they reside).

The multi-market model is then

$$\text{minimize} \quad \sum_i \sum_j x_{ij} C_{ij}$$

$$\text{subject to} \quad \sum_i x_{ij} \geq D_j, \quad \text{for each } j, \quad j = 1, \ldots, M;$$

$$\sum_j x_{ij} \leq K_i, \quad \text{for each } i, \quad i = 1, \ldots, N;$$

$$\sum_{i \in H_k} x_{ij} \geq D_{jk}, \quad \text{for each } j \text{ and } k;$$

$$x_{ij} \geq 0, \quad \text{for all } i \text{ and } j.$$

We will call this Model M and denote the optimal values of the decision variables as x^{*M}_{ij}.

The dual of model M is

$$\text{maximize} \quad \sum_j r_j D_j + \sum_j \sum_k r_{jk} D_{jk} - \sum_i s_i K_i$$

subject to $r_j + r_{jk} - s_i \le C_{ij}$, where $k = A_i$, for all i, j;

$r_j \ge 0$, for all j;

$s_i \ge 0$, for all i;

$r_{jk} \ge 0$, where $k = A_i$, for all i, j.

By complementary slackness, if $x_{ij}^{*M} > 0$, then $r_j^{*M} + r_{jk}^{*M} - C_{ij} = s_i^{*M}$. To create economically efficient incentives similar to those obtained for the single market models, the price for patient-type j would differ by market area, i.e., $p_{jk}^{*M} = r_j^{*M} + r_{jk}^{*M} + c$ (where the constant c is chosen as before to achieve budget-neutrality). Again, for every patient type j that society would want treated at hospital i ($x_{ij}^{*M} > 0$), hospital i would earn a per case profit of s_i^{*M}. Hospital i would have no incentive to treat other patient types because they would be no more profitable.

However, as noted above, we only consider pricing policies that pay a uniform price for patient-type j regardless of market area. It is unlikely that such a uniform pricing policy will create economically efficient incentives in all market areas. If the price paid for patient-type j is r_j^{*M}, then the "error" in price compared with the price that should be paid to create perfect incentives in market area k is r_{jk}^{*M}. The size of this error, which is the marginal cost of treating one more patient of type j in market area k, is a measure of the inefficiency with which area k treats patient-type j. (Since $D_j > \sum_k D_{jk}$, for every patient-type j, there is at least one market area k where $\sum x_{ij}^{*M} > D_{jk}$. Therefore, by complementary slackness, for every patient-type j, there is at least one market area k where $r_{jk} = 0$. It is in these areas that regional treatment centers for patient-type j might be encouraged.) The total pricing error weighted by the number of cases of every patient-type in every market area is

$$E = \sum_j \sum_k r_{jk}^{*M}\left(\sum_{i \in H_k} x_{ij}^{*M}\right) = \sum_j \sum_k r_{jk}^{*M} D_{jk}.$$

We consider two approaches for extending the definition of the disincentive index to the multi-market problem. Under the first approach, the profit-maximization problem for each individual hospital is formulated in the same way as in the single market case and $OC_j(p)$ is calculated similarly (using x_{ij}^{*M} in place of x_{ij}^{*S}). Thus, the disincentive index for the multi-market model is

$$V_M(p) = \sum_i \sum_j O_{ij}(p) x_{ij}^{*M} \Big/ T. \tag{5.1}$$

This first approach assumes that a hospital, in determining its most profitable case mix, ignores geographical proximity in its estimate of demand by patient-type.

An alternate approach is to formulate the profit-maximization problem for the hospital assuming that it can only attract patients from its own market area. This creates the problem of determining what number to use for market-specific demand by patient-type. One possibility is to use the number of patients of each type *currently* treated by all hospitals that compete in the area regardless where the patients reside, and this is what we use in the empirical study in §6. Another possibility is to use the number of patients of each type that will be treated by hospitals that compete in the area under the optimum solution (i.e., $\sum_{i \in H_k} x_{ij}^{*M}$). For the latter possibility to be feasible, we would have to reintroduce our benign regulator, who would suggest quotas on each patient-type for the market area. Thus, hospitals would know the types of inter-market transfers that would be encouraged by the pricing system, and would use this information in determining their most profitable case mix.

In Appendix B, we prove that if hospitals ignore market areas in their individual optimization problem, the equilibrium prices from the *single* market area model (Model S) create the best possible incentives by minimizing $V_M(p)$. Further, we show that the minimum value of $V_M(p)$ will not exceed the pricing error, E.

When market-specific demand is used in each hospital's profit maximization problems, it is not necessarily true that the equilibrium prices from the single market area model are best in the sense of minimizing $V_M(p)$. Nevertheless, we can show (see Appendix C) that equilibrium prices derived from the multi-market area model (Model M) have one important advantage: they eliminate inter-market opportunity costs when the optimal societal allocation is achieved. That is, when prices are determined from the dual of Model M, hospitals in a given market area will collectively have no incentive to exceed society's optimal allocation of patients to that area.

It is an empirical question in any particular problem how much improvement (i.e., reduction in the disincentive index), if any, is possible by using equilibrium prices, whether derived from Model S or Model M, compared to average-cost prices. It is this question to which we turn in the next section.

6. An Empirical Study

To examine the potential of equilibrium prices to improve economic incentives, we consider the problem of establishing prices for an all-payor prospective payment system for 63 hospitals that compete in eastern Massachusetts. By an all-payor system, we mean one that includes patients whether the third party payor is Medicare, Medicaid, Blue Cross, or a private insurance company.

6.1. Defining Market Areas

Our reason for including multiple market areas is to ensure that hospitals that find different patient-types economically attractive are in reasonable geographical proximity to patients in those patient-types. As noted in §5, the definition of what is "reasonable" is arbitrary and will depend on the particular situation. To develop market areas for this study, we began with hospital service areas formed by The Health Planning Council for Greater Boston for the purpose of small area variation studies of hospitalization rates in Massachusetts (Barnes et al. 1985). These areas were defined based largely on geographical proximity and patient origin data. We combined their market areas into larger areas within which travel time would be under approximately 30 minutes. The result was 4 market areas: Boston, where 14 hospitals compete; the South, where 14 hospitals compete; the West, where 12 hospitals compete; and the North, where 23 hospitals compete. In 1984, the year's data that were used in this study, between 65% and 80% of patients in each area were discharged from hospitals that competed in that area.

6.2. Patient-types

As noted, we refer to the groups of patients resulting from the aggregation of DRGs into units about which management decisions might reasonably be made as patient-types. In our example, we defined 66 patient-types, consisting of 17 general clinical subspecialties and 49 procedures or specific product lines about which separate decisions could be made. A list of DRGs comprising each patient-type is available from the authors on request.

6.3. Cost Estimation

The average cost for each hospital to treat each patient-type was estimated from two sets of data that all hospitals in Massachusetts are required to submit to the Massachusetts Rate Setting Commission. Using 1984 data from the "403 Form," the ratio of costs to charges in each of 40 revenue centers was determined for each hospital. From the "Case Mix Tape," charges for each patient in each revenue center in each hospital were available. By multiplying patient charges by the ratio of costs to charges applicable to that hospital and revenue center, and then averaging over patients, the average cost for each patient-type in each hospital was estimated. This is similar to the method that was used in the PPS to convert charges to costs.

6.4. Quality of Care Considerations

Concerns about quality of care can be incorporated by adding constraints to the model that prohibit certain hospitals from being allocated those patient-types about which there might be questions concerning quality. In our example, we added constraints to prohibit any allocation of patient-types to hospitals where the hospital had not treated at least 10 patients in 1984.

6.5. Results

Table 1 shows the value of the disincentive index for average-cost prices, equilibrium prices derived from the single market problem (Model S), and equilibrium prices derived from the multi-market model (Model M) for

Table 1 The Disincentive Index for the Multiple Market Situation Under Different Pricing Systems

	No Market Area Constraints in Model h_i	Market Area Constraints in Model h_i
Average-cost Prices	\$2671	\$991
Equilibrium Prices Model M	518	216
Equilibrium Prices Model S	483	180

two formulations of the individual hospital's optimization problem, one that ignores market areas and one that considers market areas. We prove in Appendix B that if hospitals do not consider market boundaries in their individual optimization problem, equilibrium prices derived from Model *S* are optimal. In this study, they are also the best when hospitals do consider market areas. The most notable result is that, regardless of how hospitals formulate their individual optimization problem, economic incentives are greatly improved (a minimum of a 78% reduction in the disincentive index) by equilibrium prices.

Table 2 shows the substantial reduction (89% or more) in the disincentive index under equilibrium pricing for those patient-types that comprise 5% or more of the discharges. Table 3 shows the ratio of the equilibrium price to the average-cost price for the major clinical subspecialty patient-types and for several selected procedure patient-types. This table illustrates a more general observation—equilibrium prices are higher than average cost for many of the major clinical subspecialty patient-types and lower for many of the more specialized treatments within clinical subspecialties. Thus, for many specific procedures, average-cost prices are higher than necessary and create incentives to devote more capacity than needed to these modes of treatment.

It is important to note that dollar improvements in the disincentive index are not related to cost reductions in the health care system. The actual dollar savings in health care will come from two sources: one, reduced cost because of less overexpansion; and two, as prices are adjusted yearly, cost savings due to an improved allocation of patients to those hospitals that treat them more efficiently. Overall, total costs associated with the optimum allocation of patients to hospitals is about 25% lower than current costs, under the assumption that all patients can be motivated to change hospitals. However, the actual cost saving associated with reallocation will depend on the percentage of patients that would actually change hospitals. For example, if only 10% of patients change hospitals, overall costs would be reduced 2.5%.

We examined the implications of alternative payment systems on total reimbursement to the 63 hospitals in our study given their current patient mix. A change from cost-based reimbursement to average-cost pricing (the type of change implemented when the PPS went into effect) would change payment by more than 10% to 33 of the 63 hospitals in our sample. A change from average-cost pricing to equilibrium pricing would change payment by more than 10% in only 9 of the 63 hospitals; in 42 of the 63 hospitals the change in payment would be 5% or less.

Table 2 Disincentive Index by Major Clinical Subspecialty Patient-types[1]

Patient-type	With Average-cost Prices	With Equilibrium Prices (Model S)
Cardiology (9.5)[2]	$1612	$184
Gastroenterology (6.4)	1411	109
General Surgery (9.9)	1147	107
Neonatology (8.4)	497	51
Obstetrics (10.1)	1228	45
Orthopedics (7.5)	1188	82
Pulmonary (5.8)	1346	124

[1] The disincentive index is calculated for the situation with multiple market areas under the specification that hospitals recognize market area constraints in formulating their individual optimization problem. Clinical subspecialty patient-types listed are those that include over 5% of the discharges.

[2] Percent of discharges of that patient-type.

Table 3 Ratio of Equilibrium Price (Model S) to Average Cost Price

Major Subspecialty Patient-types	Ratio
Cardiology	1.14
Gastroenterology	1.29
General surgery	0.96
Neonatology	1.51
Obstetrics	1.39
Orthopedics	1.07
Pulmonary	1.16
<u>Selected Procedure Patient-types</u>	
Cardiac Pacemaker Implant	0.70
Cardiac Catheterization	0.84
Obesity Procedures	0.75
Pancreas, Liver, and Shunt Procedures	0.71
Extracranial Vascular Procedures	0.66
Craniotomy	0.61
Coronary Bypass	0.90
Major Reconstructive Vascular Procedures	0.56

7. Discussion

The underlying rationale for our approach is that prices in a prospective payment system should be established in a way that motivates desirable behavior; and that alternative sets of prices should be evaluated according to the extent to which they motivate such behavior. Our focus has been on economic criteria, and thus we define desirable behavior as economically efficient behavior. However, other criteria could be used. For example, if some generally accepted index of quality of care existed, prices could be determined from a model that maximized quality of care (subject to a budgetary constraint) rather than minimizing costs. Also, one might argue that economic considerations (i.e., prices) play a small part in motivating decisions about case-mix changes. Even so, it is desirable to set prices in a way that rewards (or at least does not penalize) decisions that improve overall system performance.

It is important to distinguish our general approach, which derives prices from the dual of an optimal allocation model, from the particular model formulations that we used in this study. Though we feel our formulations are reasonable, there are many alternatives. As in any modeling exercise, alternative model formulations and results should be reviewed with health planners, policy analysts and hospital representatives, among others, as part of an interactive process to develop more appropriate models to help guide price setting.

In what follows, we discuss several aspects of our particular formulation.

7.1. Average Costs and the Assumption of Linearity

We used average costs as the basis for making profitability determinations. In theory, management should base decisions on variable costs and on anticipated changes in variable costs as a function of volume. However, most hospitals are unable to estimate either disease-specific variable costs or volume-related changes in cost, and third party payors are unable to collect this type of data. Thus, in current prospective payment systems, prices are based on average costs, and cost-volume relationships are not explicitly incorporated. This is the reason we used average costs in our analysis and assumed that profit increased linearly with volume. If variable costs per patient within a DRG were known and if DRG prices were set based on differences in average variable costs, our approach would be just as useful, though cost would refer to average variable cost rather than average total cost.

For smaller DRGs, cost estimates at individual hospitals are unlikely to be reliable. For these DRGs, one might want to estimate costs by averaging over several years of data. Also, to better understand the sensitivity of pricing results to possible errors in cost estimation, it would be of value to rerun the model after slightly perturbing costs.

7.2. Patient-types

Many DRGs, for example, those with the same principal diagnosis but distinguished only by whether or not complications or comorbidities existed, are not a category of patients about which a hospital could easily make a case mix decision. Similarly, a hospital that performed general surgery would likely perform both inguinal and femoral hernia procedures (DRGs 161-162) and other hernia procedures (DRGs 159-160). Thus, as discussed, for our purposes it is necessary to aggregate DRGs into categories about which decisions can be made. Though the specific way in which we aggregated DRGs to define patient-types is subject to discussion, we believe that it is important to define some aggregation of patients into decision-making units. The focus on product line management in the health care management literature (Charns and Smith 1989) is consistent with our focus on patient-types as decision-making units.

However, our use of patient-types does raise a problem. Since patient-types are aggregations of DRGs, a hospital's cost per patient-type can be thought of as the weighted average of the hospital's average cost per DRG, for those DRGs comprising the patient-type. The weights are the proportion of the hospital's patient-type caseload in each DRG. Thus, a hospital with a low cost per patient-type may be efficient, or it may be that the hospital treats more patients than average in the less expensive DRGs within the patient-type.

Since by definition we have assumed that a hospital cannot make case mix decisions about the number of patients in specific DRGs within a patient-type, the mix of patients at the hospital under a reallocation may not reflect the hospital's current DRG mix within the patient-type. Rather it might more closely reflect the average DRG mix of all patients of that type in the area. This suggests the hospital's current cost per patient-type may not be the best cost to use in deriving equilibrium prices. An alternative approach would be to use each hospital's cost per patient-type calculated as follows: the sum over those DRGs comprising the patient-type of the hospital's DRG-specific cost weighted by the proportion *of the area's* patient-type caseload in the DRG.

7.3. Severity of Illness

A variety of concerns have been raised about the failure of DRGs to adequately measure severity (e.g., Horn et al. 1985). If a severity measure were added to the PPS, the same approach just described to adjust for DRG differences within patient-type could be used to adjust for severity differences. That is, a hospital's cost per patient-type would be calculated as the sum, over those DRGs comprising the patient-type, of its DRG-severity class cost times the proportion of the patient-type caseload in the area in that class.

In the absence of severity adjustment, it is unclear if the possible failure of DRGs to adequately measure severity creates more of a problem under average-cost pricing or equilibrium pricing. Under average-cost pricing, a hospital that has a more severe case mix than average of a certain patient-type would be inadequately reimbursed. Under equilibrium pricing, the price for a patient-type is determined by the costs at those hospitals that treat patients of that type more efficiently. To the extent that some hospitals look more efficient because they treat less severe patients, the price for that patient-type will be incorrectly set at too low a level. Since it is really only relative prices that matter in a budget-neutral system, the question in regard to equilibrium prices is whether the biases in pricing caused by unmeasured severity differences occur disproportionately for different patient-types. And the real question, both in regard to average-cost pricing and equilibrium pricing, is whether biases that exist in pricing specific DRGs or patient-types cancel out when total reimbursement to the hospital is considered. Without an accepted system to measure severity, these questions cannot be answered.

7.4. Specifying Caseloads for Reallocation

In describing the model, we specified that the percentage of each hospital's current caseload available for case-mix changes and the percentage of patients of each type that could be motivated to change their historical pattern of hospital usage are the *same* for all hospitals and patient-types. Specifying that each hospital has available the same percentage of its current caseload for case-mix changes results in equilibrium prices that enable each hospital to compete successfully at its current size. Though the mix of patients that a hospital will be able to attract in a competitive environment might be different from its current mix, each hospital is provided the same opportunity to maintain its current overall market share. Thus, our system should appeal to hospitals. However, in the same way that DRG prices are recalculated periodically in the PPS, equilibrium prices would be similarly recalculated. At the time of recalculation, the new prices would reflect the overall success of each hospital in increasing its size or in improving its efficiency.

Our model could also reflect a policy decision to put pressure on the less efficient hospitals to reduce costs in order to survive and to encourage the more efficient hospitals to expand. This decision could be incorporated in the formulation by making the capacity available for case-mix changes a variable constrained to be within some range of the "fixed percentage" capacity. Likewise, one might decide that certain types of patients will be less susceptible to competitive pressures, and thus could specify that a lower proportion of these types could be motivated to change hospitals. Such a specification might be reasonable for patient-types characterized by complicated surgical procedures, for which equipment, technology, and experience are only available at limited institutions. However, we feel a better approach is to add constraints that limit the hospitals that can compete for these patients, rather than to limit the number of patients of these types available for reallocation.

7.5. The Empirical Study

It is of practical significance that in the empirical study, equilibrium prices derived from the single market area problem are the best of the alternatives considered for the multi-market problem even when market constraints are considered in the individual optimization problem.

Though we have not tested this, and though more research is clearly warranted, we suspect that as long as relatively large areas are defined, equilibrium prices from the single market problem are likely to be at least close to optimal. Given the lack of any clear basis for defining market areas, it would be useful to know that not much is sacrificed by ignoring market considerations in establishing prices.

We incorporated quality of care considerations by excluding hospitals from increased specialization in the treatment of certain patient-types unless they currently treated some minimum number of patients. A much more realistic basis for limiting increased specialization might be based on case-mix adjusted mortality rates, e.g., as calculated by the Health Care Financing Administration (Sullivan and Hays 1989). In either case, we are able to constrain the model so that we preclude allocating certain patient-types to hospitals with low costs but poor quality of care.

In our example, we found prices for an all-payor system. However, most reimbursement systems are payor-specific. In this situation, conflicting incentives created by different payor's payment systems will reduce the potential usefulness of any improvement in incentives associated with a single payor's approach. For this reason, equilibrium pricing would have the greatest potential in a statewide all-payor system. However, it is not necessary that prices be the same for each payor. For example, payor could be a factor in defining patient-types. What is important is that a single entity establish prices for the different payors in a way that overall incentives are as good as possible.

In our example, the optimum allocation is degenerate. This implies that the budget-neutral dual prices are not unique (Rubin and Wagner 1990). Though this problem might not arise in a more complete application with more constraints, there is nothing to preclude it. Thus, there may well be a large number of sets of prices that are equally good at encouraging economically efficient behavior. This emphasizes the importance of developing other criteria that can be used to evaluate sets of prices.

7.6. Conclusions

In conclusion, we wish to emphasize our main thesis: given a formulation of the patient allocation problem as a linear programming model, it is possible to

(1) determine an optimal allocation of patients to hospitals;

(2) determine a set of prices that creates incentives for hospitals to act in a way consistent with this optimal allocation;

(3) compare alternative sets of prices in terms of the extent to which they create incentives for hospitals to act in a way consistent with the optimal allocation.

There are other uses for information generated by the model. In a state with a health planning program that analyzes the amount and location of new technologies and bed expansion, the optimum allocation suggests for each patient-type which hospitals should be encouraged to expand. Also, increasingly, third party payors and managed health care programs are negotiating payment rates directly with hospitals. The dual prices from a model formulated from a single payor's perspective could provide useful guidelines in the rate negotiating process.

In summary, our contribution goes beyond proposing and justifying a specific alternative to average-cost pricing. We have, in fact, proposed a systematic and rigorous methodology for converting assumptions and goals about health care and hospitals into a pricing policy. With the huge sums of money at stake in the health care system, we believe that some such methodology is needed to guide pricing policy.[1,2]

[1] This project was supported in part by a grant from The Robert Wood Johnson Foundation. We would particularly like to acknowledge the interest and encouragement of Jeffrey Merrill at The Robert Wood Johnson Foundation. Lisa Iezzoni was of invaluable assistance in defining patient-types. Very helpful comments were received on an earlier draft of this paper from Joseph Restuccia, Thomas McGuire and Randall Ellis.

[2] Appendices are available upon request from TIMS Editorial Offices, 290 Westminster Street, Providence, Rhode Island 02903.

References

Barnes, B. A., E. O'Brien, C. Comstock, D. G. D'Arpa, and C. L. Donahue, "Report on Variation in Rates of Utilization of Surgical Services in the Commonwealth of Massachusetts," *J. of the American Medical Association*, 254, 3 (1985), 371–375.

Charns, M. P. and L. J. Smith, "Product Line Management and Continuum of Care," *Health Matrix*, 7, 1 (1989), 40–49.

Cotterill, P., J. Bobula, and R. Connerton, "Comparison of Alternative Weights for Diagnosis-Related Groups," *Health Care Financing Review*, 7, 3 (1986), 37–51.

Dantzig, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.

Fetter, R. B., Y. Shin, J. L. Freeman, R. F. Averill, and J. D. Thompson, "Case Mix Definition by Diagnosis-Related Groups," *Medical Care*, 18, 2 (1980), 1–53.

——, "Diagnosis Related Groups: Understanding Hospital Performance," *Interfaces*, 21, 1 (1991), 6–26.

Gale, D., *The Theory of Linear Economic Models*, McGraw Hill, New York, 1960.

Garnick, D. W., H. S. Luft, J. C. Robinson, and J. Tetreault, "Appropriate Measures of Hospital Market Areas," *Health Services Research*, 22, 1 (1987), 69–89.

Horn, S. D., G. Bulkley, P. D. Sharkey, A. F. Chambers, R. A. Horn, and C. J. Schram, "Interhospital Differences in Severity of Illness: Problems for Prospective Payment Based on Diagnosis-Related Groups," *New England J. Medicine*, 313, 1 (1985), 20–24.

Koopmans, T. C., "Analysis of Production as an Efficient Combination of Activities," in *Activity Analysis of Production and Allocation*, T. C. Koopmans (Ed.), John Wiley & Sons, New York, 1951.

Price, K. F., "Pricing Medicare's Diagnosis-Related Groups: Charges Versus Estimated Costs," *Health Care Financing Review*, 11, 1 (1989), 79–90.

Rubin, D. S. and H. M. Wagner, "Shadow Prices: Tips and Traps for Managers and Instructors," *Interfaces*, 20, 4 (1990), 150–157.

Schweiker, R. S., *Report to Congress: Hospital Prospective Payment for Medicare*, U.S. Department of Health and Human Services, Washington, D.C., 1982.

Sullivan, L. W. and L. B. Hays, *Medicare Hospital Mortality Information*, 1986, 1987, 1988, Health Care Financing Administration, U.S. Department of Health and Human Services, Washington, D.C., 1989. HCFA Publication 00701.

Wennberg, J. E., B. A. Barnes, and M. Zubkoff, "Professional Uncertainty and the Problem of Supplier-Induced Demand," *Social Science and Medicine*, 16 (1982), 811–824.

Accepted by Arnold Barnett, former Departmental Editor.