Payer Competition and Cost Shifting in Health Care

JACOB GLAZER

Department of Economics
Boston University
and the School of Management
Tel-Aviv University

THOMAS G. MCGUIRE

Department of Economics
Boston University
Boston, MA 02215

This paper studies a model in which two payers contract with one hospital. True costs per patient are not a possible basis for payment, and contracts can only be written on the basis of allocated cost. Payers choose a contract that is fully prospective or fully based on cost allocation, or a payment scheme that would give some weight to each of these two. We characterize the payers' equilibrium contracts and show how in equilibrium hospital input decisions are distorted by the payers' incentives to engage in cost shifting. Two cost-shifting incentives work in opposite directions, and equilibrium can be characterized by too little or too much care relative to the socially efficient level.

1. Introduction

Private insurers compete for enrollees by setting premiums and the terms of coverage. They also compete by buying services from the same set of hospitals and other health care providers. The Clinton administration would rely on "managed competition" to regulate payer competition, with almost all public debate to date, however, being concerned with payer product-side competition, involving such issues as universal coverage, mandated benefits, community rating,

This paper was prepared for the conference "The Industrial Organization of Health Care," which was held at Boston University's Osgood Hill Conference Center, North Andover, Massachusetts, September 19–21, 1993. Financial support from the Management Science Group, Veterans Administration (Glazer) and the National Institute of Mental Health (McGuire) is gratefully acknowledged. Albert Ma, Dan Spulber, and an anonymous referee contributed many helpful suggestions.

Journal of Economics & Management Strategy, Volume 3, Number 1, Spring 1994, 71–92

and elimination of preexisting-condition exclusions (Enthoven and Kronick, 1991; Clinton, 1993).

Payer competition in hospital contracting also deserves attention. Competition by payers may, of course, encourage efficient behavior by hospitals. Depending on the possible forms of hospital contracts, competing payers may also, however, attempt to shift hospital costs to rivals and in so doing encourage socially wasteful activity by hospitals. Suppose, for example, that one payer at a hospital pays a share of total cost in proportion to its share of total days, and the other payer pays a fixed price per discharge. The hospital has a clear incentive to increase days for the payer using the allocation formula and decrease days to the prospective payer in order to increase hospital net revenue.[1] This example suggests that using a cost-allocation formula exposes a payer to unfavorable cost shifting. This logic is correct, and we believe largely explains the flight from cost-allocation payment systems that occurred in the United States in the 1980s following the federal Medicare program's move from cost-allocation to prospective payment in 1983.

As we show in this paper, cost-shifting incentives have another side. In a reversal of the usual logic, it is also true that one payer can use a *cost-allocation* system to exploit another payer's *prospective* payment system. By paying at least partially on the basis of allocated costs, a payer can induce the hospital to shift resources to its patients and away from patients of the payer making only a prospective payment. The prospective payment would then more than cover cost, and the payer making the cost-allocation payment would have succeeded in cost shifting onto the prospective payer. In this paper, we explore the full range of cost-shifting incentives associated with allocation systems and analyze their implications for equilibrium and welfare in a model of payers' competition.

We study a model in which two payers contract with one hospital. Each of the payers chooses how to reimburse the hospital for treating its patients. The hospital, in turn, decides whether to accept (each of) these contracts and also decides on the level of care (length of stay, in our model) to be provided to each of the patients.

The key feature of our model is that we restrict the set of contracts that the payers can offer the hospital. First, we assume that

1. As Glaser puts it in his book, *Paying the Hospital* (p. 83), "Every American hospital must employ a clever chief financial officer who can distribute profits and costs among payers, so that they can detect and protest as little as possible. Most of the hospital's profits and much of the costs not accepted by the more restrictive payers must be charged surreptitiously to the others." See Frankfurt (1993) for a more recent discussion.

payers cannot simply "dictate" to the hospital their patients' length of stay. In reality, a payer contracts with many hospitals, each serving hundreds of the payers' patients. It is simply infeasible for a payer to dictate hospital production. The second assumption we make is that the hospital cannot be reimbursed on the basis of actual cost per patient. In spite of the term *cost-based* being applied to some hospital payment schemes, no real-world method for hospital payment uses *actual costs* per case for the simple reason that actual costs are not observable to payers. The amount of nursing time, for example, devoted to care of a patient is not routinely recorded in hospitals. Much of the time spent by nurses and other hospital workers may not even reasonably be assigned to particular patients, even if it were recorded and observed—similarly with many elements of capital. A payer may be able to observe that some piece of equipment was used but may not know the intensity of that use.

Some inputs are of course readily observable and potentially contractible. The most easily observed input into hospital care is the number of overnights the patient spends in the hospital, known as the patient's "length of stay" (LOS). The admission and discharge dates are routinely recorded on all hospital claims for payment, even those claims to payers using a prospective payment system. It is no coincidence that LOS figures very prominently in actual payment systems.[2] (Indeed, some payers make exclusive reliance on LOS, employing a per diem payment system under which the hospital is paid a constant price per day.)

In this paper we recognize that true costs per patient are not a possible basis for payment and study "cost-based" payment where cost is allocated cost. Cost allocation methods assign to a payer or a discharge a share of the hospital's total costs based on the utilization of one or a small set of the observable resources used by the hospital. A common cost-allocation method assigns a share of total hospital room-and-board costs according to the payer's share of total days used.[3]

2. Other hospital inputs are also observable, such as the use of certain facilities (operating room, intensive care unit) and equipment (diagnostic tests), or the receipt of certain special services (respiratory therapy); some payers pay on the basis of the "cost" or a price for these ancillary services.

3. Another class of payment methods employs a price for one or a few observable hospital inputs. A per diem payment system in which the payer pays the hospital a fixed price per day is an example of a payment based on the use of only one input. The price per day must cover costs for the priced and the unpriced resources. Price-based payments for a subset of hospital inputs are a common form of hospital payment. See Tables I–III.

This second assumption is an important departure from the current literature on hospital reimbursement. A number of papers compare cost-based payment, prospective payment, and a mixed payment system with a prospective and cost-based component, assuming actual patient costs are contractible. The main finding in this literature is that when a hospital decision maker weighs both patient benefit and hospital profit, a mixed system is able to induce the hospital to provide the first-best outcome[4] (see, e.g., Ellis and McGuire, 1986, 1993; Goodall, 1990; Ma, 1994; Pope, 1989). This, and other conclusions from this literature, are open to question when costs for purposes of payment are allocated costs, not actual costs. As we will show, allocated costs can create quite different incentives because of the potential for cost shifting.

When cost per patient is not contractible, three bases for payment can be identified:

1. *Prospective payment:* The payer can simply set a price per discharge, requiring no observation of the value of any input or cost.
2. The payer (or the hospital) can set a price for contractible inputs, relying on a margin above cost to cover the hospital's costs for noncontractible inputs. The most common form of price-based payment is a per diem system whereby a payer pays a fixed price per day of a patient's stay, and this price is expected to cover all costs.
3. *Cost allocation:* The payer can approximate a cost-based payment system by a formula using allocated costs. The one we will consider is where the payer reimburses the hospital for a share of allocated costs.

These three methods capture the main features of observed payment systems. Obviously, a payer does not necessarily have to use only one of these bases and might actually prefer a combination of them. In fact, as the literature on hospital reimbursement mentioned earlier has shown, when cost per patient is observable, a mixture of prospective and cost-based reimbursement is often desirable.

In this paper we do not study per diem contracts and allow

4. There are several arguments for a mixed system. First, payers can neither contract on cost per discharge for each patient nor rely on competition to maximize patient surplus for a given price per discharge. Paying a share of costs mitigates a hospital's incentives to limit the supply of care in comparison with a fully prospective payment. Second, payers cannot accurately assign a "price" to every patient/disease combination. By sharing costs with the hospital, the payer reduces a hospital's incentive to avoid patients with a prospective payment set too low and compete for patients with a price set too high. Third, a mixed system reduces financial risk to the hospital. See Ellis and McGuire (1993) for elaboration.

the payers only to choose a contract that would be a combination of prospective and cost allocation.[5] They can choose a contract that is fully prospective or fully based on cost allocation, or a payment scheme that would give some weight to each of these two.

Previous analyses of hospital "cost shifting" have used this term in a different sense and been concerned primarily with hospital price-setting behavior (Dranove, 1988; Frank and Salkever, 1991; Foster, 1985; Hay, 1983; Sloan and Becker, 1984).[6] Sloan and Becker (1984), for example, define cost shifting as occurring when a reduction in the government payment rate increases the price charged to private-pay patients. As hospital prices become increasingly subject to regulation, hospital price setting diminishes in importance. Cost shifting when *payers* set terms of payment and hospitals make input decisions is the more relevant form for an era of payer competition.

Our main purpose is to characterize the payers' equilibrium contracts and in particular to see whether in equilibrium hospital input decisions are distorted by the payers' incentives to engage in cost shifting. It is easy to show in our model that if cost per patient were contractible, payers' competition would lead to the socially efficient level of care. Limiting contracts to realistic forms interferes with this result. In general, we find that the equilibrium when payers compete is not efficient. More importantly, we show that there are two cost-shifting incentives to payers that work in opposite directions, and equilibrium can be characterized by too little or too much care relative to the socially efficient level. The magnitude of the noncontractible cost in relation to the contractible input determines the importance and the direction of the distortion. The larger the share of the noncontractible costs, the more likely it is that too little care will be provided in equilibrium. We show also that under some conditions, equilibrium is such that all payers choose prospective payment, and in this case, there is no other set of contracts that improve social welfare.

We cast this problem as one of common agency.[7] The payers are

5. Per diem systems are fairly common as a basis for payment, as Tables I–III indicate. A central feature of a per diem system is that it may distort the choice of production by paying a price above marginal cost for some inputs in order to cover the costs of all inputs. Per diem systems need to be studied in a model more complex than ours that allows for substitution between inputs.

6. Ma and McGuire (1993) analyze a model in which payers set per discharge prices, and the government payer may pay a share of fixed (common) costs in the form of pass-through payments. The government player can pay less than its share of fixed costs by exploiting its first-mover advantage.

7. The common agency problem was introduced and rigorously studied by Bernheim and Whinston (1986). They obtained some general results on the conditions under which first-best and second-best outcomes will be realized in equilibrium. Our concern

the principals, and the hospital is the common agent. The objective function of each payer is to maximize its payoff, which is assumed to be its patient's benefit minus the payment to the hospital. The hospital will accept the payers' contracts only if they will at least cover its costs. This is where cost shifting becomes crucial. We assume that even if one payer's contract is not likely to cover its patients' costs, the hospital may still accept it if the other payer's contract is generous enough to balance the hospital's revenue.

The other decision the hospital makes is about the level of treatment (length of stay). Here we take the approach that the hospital cares not only about its profit but also about the patient's benefit, following much of the recent literature on hospital behavior. Participation of the patient's doctor in decisions about treatment (as opposed to the hospital management's decision about accepting contracts) is the main rationale for assuming that patient benefit as well as profit matter for treatment decisions. This assumption enabled Ellis and McGuire (1986) to show that when cost per patient is contractible, then even if patient's benefit is not contractible, a mix of prospective and cost-based payment may induce the hospital to provide the efficient level of care.[8]

Before we begin our formal analysis, we briefly survey the forms of hospital payment. We then describe the model and go on to characterize equilibrium behavior by payers and the hospital. A series of propositions contains our main results.

2. Methods of Payment

Commercial insurance companies, small in any hospital market, historically have paid hospitals' charges. In their "traditional business," meaning plans with no restrictions on enrollee's choice of hospital, commercial insurers still pay charges. Increasingly, however, commercial insurers pay on the basis of contracts with providers. In pre-

is somewhat more specific. We would like to address issues such as the conditions under which too much or too little of the variable input is provided by the hospitals, in equilibrium, or the conditions under which various payment systems will be offered by payers. For those purposes, we construct a more special model with features of the payers–hospital relationship.

8. Ellis and McGuire's result is in the spirit of Loeb and Magat (1979). By setting the degree of prospectiveness in the payment scheme, the "hospital" caring about a weighted average of profit and patient benefit can be made essentially to maximize social welfare. Loeb and Magat achieved a similar end by paying a public utility for consumer surplus.

ferred provider organization (PPO) plans, in which an enrollee has some but inferior coverage for providers not among the "preferred," and in health maintenance organization (HMO) plans, in which an enrollee has no coverage outside the covered providers, the commercial insurer typically extracts a price concession from a hospital in exchange for a listing among the preferred or exclusive providers. In 1990 about 75% of commercial insurers' premiums derived from traditional business in which the insurers paid charges, and the balance from PPO and HMO plans, but this mix was changing rapidly, however, away from the traditional business (Hoy and Curtis, 1991).

In the late 1970s, Blue Cross plans were split evenly between cost- and charge-based. Most cost-based plans paid a per diem with costs allocated on a per day basis (Berman and Weeks, 1982, pp. 152–153), a payment system that predominates in western European countries, where all payers or a single payer use the same basis of payment (Glaser, 1987, p. 45). In the early 1980s, state Medicaid plans made heavy use of per diem payments, calculated from a restricted definition of allowable costs (Glaser, 1987, p. 82).

Some Blue Cross plans and Medicaid programs continue to pay using allocated costs, but these methods have become much less common recently. Tables I–III describe the prevalence of various payment methods used by commercial insurers, Blue Cross Plans, and Medi-

TABLE I.
How Commercial Insurers Pay Hospitals: Percent of Premium Revenue by Payment Method: 1990

	Products (%)		
Payment Method	Traditional	PPO	HMO
DRG/Episode/Case	0	22	5
Per capita	0	0	7
Item of service 100% of charges	100	18	12
Item of service discount or fixed rate	0	35	71
Per diem	0	25	5
Total	100	100	100

Source: Hoy et al. (1991).
PPO, preferred provider organization; HMO, health maintenance organization; DRG, diagnosis-related group.

TABLE II.
How Blue Cross Plans Pay Hospitals: Number of Plans by Payment Method: 1991

	Products		
Payment Method	Traditional	PPO	HMO
DRG/Episode/Case	26	24	16
Per capita	2	0	1
Item of service			
100% of charges	5	0	1
Item of service			
Discount or fixed rate	21	10	6
Per diem cost analysis	4	1	2
Per diem other	6	9	18
Other		1	
Missing	3	6	9
Not applicable		16	14
Total	67	67	67

Data are for the 67 plans reporting for 1991, unpublished data from the Center for Health Economics and Policy Research, BlueCross BlueShield Association, an association of Independent Blue Cross and Blue Shield Plans, Chicago. We are grateful to Bob Lapp for help with these figures.
See Table I footnote for expansion of abbreviations.

TABLE III.
Medicaid Inpatient Payment Systems, by State: 1991

Payment Method	Number of States[a]
Retrospective	
Costs	4
Prospective limit on costs	
Cost to trend limit	6
Cost to peer and trend limit	1
Prospective Rates:	
Hospital level	10
Hospital level with peer groups	6
Diagnosis specific	22
Negotiated	4

Source: *Medicaid Source Book: Background Data and Analysis (1993 Update)*, prepared by the Congressional Research Service, January 1993.
[a] Arizona operates the Arizona Health Care Cost Containment System (AHCCCS) as an alternative to Medicaid and, thus, is not included in Table 3. California, Illinois, and Washington have two systems in effect and contribute to the totals in both categories. The District of Columbia is included.

caid programs in the United States.[9] Medicare, the federal program for the elderly and disabled, pays for most discharges prospectively using diagnosis-related groups (DRGs). Medicare uses a cost-allocation system known as The Tax Equity and Fiscal Responsibility Act of 1982 (TEFRA) to pay for about 400,000 discharges per year in specialty hospitals.

3. The Model

There are two payers denoted by 1 and 2. Each payer has only one patient (an assumption made in order to simplify the analysis). Both payers contract with the same hospital for treating their patient. We describe the cost-and-benefit functions and then proceed to describe the types of compensation schemes (contracts) the payers can choose.

3.1 Costs and Benefits of Treatment

We assume there are two inputs, X and F. Let x_i be the quantity of input X used to treat patient i. The other input is assumed to be fixed at a level F for all patients. Thus, patient i is treated with x_i units of X and F units of a second input. We will refer to x_i as a patient's "length of stay," intending x_i to connote a readily observable and contractible summary statistic of a set of inputs. F is intended as a summary of the other inputs used by a hospital, such as equipment and labor inputs that are not practical as a per patient basis of payment. If we assume the price of each input is one, the hospital's cost for treating patient i, therefore, is $C_i = x_i + F$. Costs of treating patient 1 are assumed to be independent of the costs of treating patient 2, so the hospital's total cost of treating both patients is

$$C = C_1 + C_2 = x_1 + x_2 + 2F.$$

Patient i benefits from treatment according to $B(x_i, F)$. We assume $B_x > 0$, $B_{xx} < 0$, and $B_x(0, F) = \infty$, with the x subscripts denoting partial derivatives. Marginal benefits start very large, ensuring that some x is efficient (and will be chosen in the model later), and decline. Because F is assumed to be fixed in our model, we will omit it from

9. The forms of contracting used by commercial insurers and Blue Cross plans can partly be explained by the nature of their business. Any payment with the exception of billed charges requires an agreement between the payer and the hospital. Commercial insurers with subscribers spread over many markets may not find it worthwhile to contract with hospitals and, therefore, more frequently pay charges. Blue Cross plans are state- or substate-based and always write hospital contracts.

the benefit function hereafter, so $B(x)$ in our model should be understood as $B(x, F)$, with the F suppressed.

Patients are assumed to be fully insured and, therefore, face no risk of health expenditures. Patients are also passive. In particular, x and the benefits from hospital care do not influence the quantity of discharges demanded from the hospital.

Let x^* be such that $B_x(x^*) = 1$. At x^*, the marginal benefit of x equals the marginal cost. x^* will be referred to as the efficient (first-best) x.

3.2 Feasible Payment Systems with Cost Allocation

A cost allocation method is a way to assign costs to a payer. The basic idea is to allocate all costs in proportion to a payer's use of a contractible input.[10] In our case, costs allocated to payer i must be in proportion to x_i. Therefore, we define allocated costs to payer i as follows:

$$\hat{C}_i(x_1, x_2) = \frac{x_i}{x_1 + x_2} \cdot ((F_1 + x_1) + (F_2 + x_2)). \tag{1}$$

Note that costs allocated to payer 1 plus costs allocated to payer 2 sum to actual total costs. Note also that when $F = 0$, that is, only one input is used in production, costs allocated to payer i equal actual costs of treatment for payer i, which are just x_i.

The family of payment systems we consider here is all possible combinations of the prospective payment amount per discharge (R), and the share of allocated costs born by the payer (r). Thus, payer i's contract is an (R_i, r_i) pair, with $R_i \geq 0$ and $0 \leq r_i \leq 1$. If allocated costs to payer i are $\hat{C}_i$, payer i pays $R_i + r_i\hat{C}_i$ to the hospital. This family includes complete prospective payment $(R, 0)$, cost-based payment $(0, 1)$ based on allocated costs, and the range of payments systems intermediate between these characterized by a prospective component and a cost-based component. Following Ellis and McGuire (1986), we will refer to any payment system with $R > 0$ and $r > 0$ as a "mixed" system.

As Tables 1–3 show, private insurance companies tend to use either completely prospective contracts or to rely on only allocated costs. The intermediary "mixed" systems are not observed, although we allow for these contracts in the game we analyze later. At the close

10. This may be done at a cost center level or at the level of the entire hospital. Allocation of total costs according to "days" has been used by Blue Cross plans and the public Medicare and Medicaid programs. See Berman and Weeks (1982) for discussion.

Stage 1:	**Payer 1** **Chooses (R_1, r_1)**	**Payer 2** **Chooses (R_2, r_2)**
Stage 2:	**Hospital Accepts or Rejects** **(R_1, r_1) and/or (R_2, r_2)**	
Stage 3:	**Hospital Chooses (x_1, x_2)**	

FIGURE 1. STAGES OF THE GAME.

of the paper, we will comment on the reasons for why mixed systems have not appeared in private insurance contracts.

One contract a payer could choose is to simply "dictate" x. This could be done by offering a contract that just covers the cost of the desired x but pays nothing if the hospital chooses any other x. As discussed in the Introduction, we regard this contract as infeasible in the context of many hospitals making a decision about many inputs to be used in connection many discharges, and, therefore, we disregard it in our model.

4. The Game and the Equilibrium

The game has three stages, which are summarized in Figure 1. In the first stage, both payers simultaneously offer a compensation scheme to the hospital. In the second stage, the hospital decides whether to accept these contracts, and, in the third stage, the hospital decides on the level of care to be provided to each of the patients. In order to solve the subgame perfect equilibrium of the three-stage game, we shall first analyze the hospital's behavior at stage 3. We shall characterize the hospital's choice of x_1 and x_2 for any combination of contracts (R_1, r_1) and (R_2, r_2). We then move backward to stage 2 and analyze the hospital's decision about which patients (contracts) to accept, taking into account the hospital's own behavior in stage 3. Finally, we move one more step backward and analyze how the payers choose (simultaneously) their contracts (R_i, r_i), taking into account the hospital's behavior in the later stages.

4.1 The Hospital's Choice of Length of Stay

At stage 3, a decision is made about the level of treatment each patient receives, x_1 and x_2, respectively. Writers on hospital behavior have taken different approaches to hospital objectives in setting treatment. In some models, hospitals choose x to maximize profits (e.g., Ma, 1994), or to maximize services per se or a function of services such as patient benefits (Rogerson, 1994). Other papers assume hospitals maximize a weighted combination of profits and a function of services, motivated in some cases by the involvement by the patient's physician in the decision-making process (Ellis and McGuire, 1986; Frank and Lave, 1989). In our model, one decision maker, which we refer to for simplicity as the hospital, chooses x to maximize a utility function with profit and patient benefit as arguments. Patient i's benefit, $B(x_i)$, has been defined previously. When payers pay according to a cost-allocation system, the hospital's profit, Π, is

$$\Pi = R_1 + R_2 + r_1\hat{C}_1(x_1, x_2) + r_2\hat{C}_2(x_1, x_2) - 2F - x_1 - x_2. \tag{2}$$

We assume the hospital chooses x_1 and x_2 to maximize a weighted sum of patient benefit and hospital profits, with the weights on patient benefit and profit being β and $1 - \beta$, respectively. Thus, the hospital's objective function, which it wishes to maximize, is

$$U = \beta[B(x_1) + B(x_2)] + (1 - \beta)\Pi(x_1, x_2). \tag{3}$$

We do not model the interactions between doctors, patients, and hospital management, which may be understood as leading to the β weights.[11] We simply assume here that the choice of x_i is Pareto-efficient in the (B, Π)-space.

Maximizing eq. (3) with respect to x_1 and x_2 after substituting in the expression for profit, eq. (2), we have two equations described in eq. (4), which we will refer to as the incentive compatibility constraints.

$$\beta B_x(x_i) + (1 - \beta)\left((r_i - r_j)\frac{\hat{C}_j(x_1, x_2)}{(x_1 + x_2)} + \frac{r_1x_1 + r_2x_2}{x_1 + x_2} - 1\right) = 0 \qquad \text{for } i, j = 1, 2, i \neq j. \tag{4}$$

Equation (4) shows the incentives surrounding the hospital's choice

11. This objective function could be regarded as the outcome of a bargaining process among the interested parties.

of x_i, including the incentives to engage in cost shifting. When x_i goes up, the hospital benefits because it values patient benefit (the first term). Part of the cost is borne by the hospital when the payers set r_i and r_j less than one. Because the payment systems work on cost allocation, the total hospital costs are divided between the payers in proportion to their x's. Thus, the share of marginal costs borne by the hospital is the last term in eq. (4). (This is weighted by $(1 - \beta)$). The middle term in eq. (4) leads the hospital to cost-shift. Suppose $r_i > r_j$, meaning payer i pays a higher share of costs at the margin. By increasing x_i, the hospital allocates more of costs to payer i and gains the difference in r's times the amount of cost reallocated. This term describes the hospital's incentive to allocate inputs to maximize net revenues.

Before we proceed, we would like to show that there is a level or r, call it r^*, that would induce the hospital to provide the first-base level of care x^* under the cost allocation payment scheme. Let

$$r^* = (1 - 2\beta)/(1 - \beta). \tag{5}$$

Then it can be shown that if $r_1 = r_2 = r^*$, the hospital would choose $x_1 = x_2 = x^*$. One important question to address in our model is whether competition in the presence of cost allocation leads to too high or too low r, (and x) in relation to the efficient levels.

4.2 The Hospital's Decision About Which Contracts to Accept

At stage 2, the hospital decides to accept or reject payers' contracts. We assume that the hospital's objective is to accept as many patients as possible as long as it covers its costs. Thus, if the two contracts combined are such that they cover the hospital's cost (given that the hospital chooses inputs according to eq. [4]), the hospital will accept the contracts. Suppose that payer 1 offers contract (R_1, r_1), and payer 2 offers contract (R_2, r_2). Will the hospital accept these contracts? If the hospital accepts these two contracts, it will choose x_1 and x_2 according to eq. (4). Then, we assume the hospital will accept these two contracts if:

$$R_1 + R_2 + r_1\hat{C}(x_1, x_2) + r_2\hat{C}_2(x_1, x_2) - (2F + x_1 + x_2) \geq 0. \tag{6}$$

If eq. (6) does not hold, we assume that the hospital may take just one contract, if that contract covers costs. Formally, we could introduce a new set of constraints to cover this possibility, but it is easy to show that in equilibrium, the binding constraint will always be eq. (6). We will refer to eq. (6) as the individual rationality constraint.

The importance of condition (6) is that it tells us which deviations are acceptable to the hospital. Suppose that both payers offer some contract (R, r). If one of the payers considers deviating and offering (unilaterally) another contract (R', r'), say, it must guarantee that this contract, together with the other payer offering (R, r), will satisfy condition (6). Otherwise, the hospital will not accept this new contract.

4.3 Payer Choice of Contract

At stage 1, the payers simultaneously set contracts. We assume that the payer's objective is to maximize the difference between the patient's benefit and the payment to the hospital. The idea is that this patient/payer surplus, defined by the difference just mentioned, is divided somehow between the patient and the payer, and, therefore, it is in the payer's interest to make the surplus as large as possible. Alternatively, one could assume that many payers compete for the patient, and the patient chooses the one offering the highest surplus.

Given these objective functions of the payers, denoted W_1 and W_2, respectively, one can easily show that if cost per patient were contractible, equilibrium would result in both payers offering a contract where $r_1 = r_2 = r^*$. That is, payers' competition would lead to the hospital providing the efficient level of care. We see now, however, what happens when cost per patient is not contractible.

At the subgame perfect Nash equilibrium of this game, payer i's contract, (R_i^e, r_i^e), should maximize W_i, given payer j's contract, (R_j^e, r_j^e), and given the hospital's behavior at stage 2 as given by eq. (6), and at stage 3 as given by the pair of eqs. (4). Specifically, payer i's problem can be written as follows:

$$\max_{R_i, r_i, x_1, x_2} W_i = B(x_i) - R_i - r_i[\hat{C}_i(x_1, x_2)], \tag{7}$$

subject to eqs. (4) and (6).

This is a common agency problem. Each principal (payer) maximizes its payoff subject to the other principal's contract, the agent's (the hospital) incentive compatibility constraints (4), and the agent's individual rationality constraint (6).

It is easy to show that in equilibrium, eq. (6) must hold as an equality. Therefore, we can substitute eq. (6) into payer i's objective function, and, writing out the expression for $\hat{C}_i(x_1, x_2)$, we can describe payer i's problem as

$$\underset{r_i, x_1, x_2}{\text{Max}}\ W_i = B(x_i) + [r_j x_j/(x_1 + x_2) - 1](2F + x_1 + x_2) + R_j, \tag{8}$$

subject to eq. (4).

This formulation makes clear that since the hospital makes exactly zero profit, payer i must pay the difference between total costs and payer j's payment, $R_j + r_j\hat{C}_j(x_1, x_2)$.

5. Analysis and Findings

Any pair of contracts that simultaneously solves the constrained maximization problem (8), for both $i = 1$ and $i = 2$, is a pair of equilibrium contracts. We will focus on symmetric equilibria where $R_1^e = R_2^e$, and $r_1^e = r_2^e$.

Notice that r_i does not appear in the objective function of eq. (8) and, therefore:

$$\frac{dW_i}{dr_i} = \frac{dW_i}{dx_i}\frac{dx_i}{dr_i} + \frac{dW_i}{dx_j}\frac{dx_j}{dr_i}. \tag{9}$$

Using the incentive compatibility constraint (4), we can show that if $r_1 = r_2 = r$, then $x_1 = x_2 = x$, and

$$\frac{dx_i}{dr_i} = -\frac{(1-\beta)\left(\dfrac{F}{2x} + 1\right)}{\beta\, B_{xx}} > 0 \tag{10}$$

$$\frac{dx_j}{dr_i} = \frac{(1-\beta)\dfrac{F}{2x}}{\beta\, B_{xx}} < 0. \tag{11}$$

Equations (10) and (11) describe the behavior of the hospital bearing on cost shifting. If r_i is increased leaving r_j unchanged, the hospital will shift inputs from patient j to patient i.

From the definition of W_i we know that if $r_1 = r_2 = r$ and $x_1 = x_2 = x$, then

$$\frac{dW_i}{dx_i} = B_x(x) - r\frac{F}{2x} - 1, \tag{12}$$

and

$$\frac{dW_i}{dx_j} = r\frac{F}{2x} + r - 1. \tag{13}$$

Now, substituting eqs. (10)–(13) into eq. (9), we get that when $r_1 = r_2 = r$,

$$\frac{dW_i}{dr_i} = \frac{(1-\beta)}{\beta B_{xx}}\left(\left(B_x - r\frac{F}{2x} - 1\right)\left(-\frac{F}{2x} - 1\right) + \left(r\frac{F}{2x} + r - 1\right)\left(\frac{F}{2x}\right)\right) \tag{14}$$

for $i = 1, 2$.

The equilibrium of this game can either be at the corner, where $r_1 = r_2 = r = 0$, or in the interior, where $r_1 = r_2 = r > 0$. In order for an equilibrium to be at the corner, it must be that the derivative (14) evaluated at $r_1 = r_2 = 0$ is less than or equal to zero, implying that each payer i would be worse off if r_i were increased away from zero. Substituting $r = 0$ into eq. (14), and given that $(1 - \beta)/\beta B_{xx} < 0$, the condition for a corner equilibrium is:

$$(B_x(x^o) - 1)\left(\frac{F}{2x^o} + 1\right) + \frac{F}{2x^o} < 0, \tag{15}$$

where x^o is the hospital's choice of x when $r_1 = r_2 = 0$.

Using the incentive compatibility constraint (4), we know that at $r_1 = r_2 = 0$, x^o is given by

$$B_x(x^o) = (1 - \beta)/\beta. \tag{16}$$

We can now state the following proposition:

PROPOSITION 1: *At the Nash equilibrium, $r_1 = r_2 = 0$ if and only if:*

$$\beta \geq \frac{F + 2x^o}{F + 4x^o} \equiv \bar{\beta} \tag{17}$$

where x^o is given by eq. (16).

Proof. The proof is obtained by substituting eq. (16) into eq. (15). □

Corollary. If $\beta < \frac{1}{2}$, payers will not pay only prospectively in equilibrium.

As can be seen from Proposition 1, when $F = 0$, equilibrium will be at a corner when $\beta \geq \frac{1}{2}$. That is, when $\beta \geq \frac{1}{2}$, payers choose to pay only prospectively, setting $r_i = 0$. However, when F is bigger than zero, then even when the hospital cares equally about the patient's benefit and its own profit (i.e., $\beta = \frac{1}{2}$), the payers will choose to have some positive retrospective component in their contract. Notice that, in contrast, if all of per patient costs were contractible, then with $\beta \geq \frac{1}{2}$, payers would pay only prospectively.

We now consider the incentives to payers when the parameters of the model imply that the socially best choice of contract is nearly fully prospective, (i.e., r is near zero), in order to show that a cost allocation system can cause payers to choose r "too high." Suppose $\beta = \frac{1}{2}$. From eq. (5) we know that the optimal contract is a fully prospective contract with $r^* = 0$. We also know from Proposition 1 that this is not an equilibrium when $F > 0$. Consider the incentives to payer i to deviate from a contract $(R_i, 0)$ when payer j's contract is $(R_j, 0)$, with $R_i = R_j = F + x_i = F + x_j$, and assuming the hospital's (4) and (6) constraints are satisfied. Assume $F > 0$, so that cost allocation is an issue. If payer i raises r_i slightly to a positive value, the hospital changes x_i and x_j according to eqs. (10) and (11). x_i rises, and x_j falls because of the incentives to the hospital associated with a cost allocation system. Since r_i has gone up, the hospital seeks to allocate more of the noncontractible costs to payer i. When x_j falls, payer j is put in the position of paying for more than it gets, that is, now $R_j > F + x_j$. Payer i exploits payer j's overpayment by reducing R_i so that eq. (6) remains satisfied as an equality. Payer i does not bear the full extra cost of the x_i induced by the rise in r_i and so has an incentive to raise r_i above the optimal level, which in this example is zero.

Both payers of course share the same incentive. In equilibrium, neither payer will succeed in cost shifting onto the other, although both will be trying by setting $r > 0$. This equilibrium will be inefficient because the quantity of the variable input x will exceed the optimal level for both payers.

We can now study the interior equilibrium where $r_1 = r_2 = r > 0$ and see how another effect of cost-allocation systems becomes evident. At the interior equilibrium, it must be that

$$\frac{dW_i}{dr_i} = 0. \tag{18}$$

Let r^e be the equilibrium r, and let x^e denote the equilibrium x at the interior solution. Given that $(1 - \beta)/\beta B_{xx} \neq 0$, condition (18) together with eq. (14) implies that at an interior equilibrium

$$B_x(x^e) - 1 - r^e(F/x^e) + F/(F + 2x^e) = 0. \tag{19}$$

From the incentive compatibility constraints (4), we also know that

$$r^e = 1 - \beta B_x(x^e)/(1 - \beta). \tag{20}$$

The first observation to make is that when $F = 0$, $B_x(x^e) = 1$, and, hence, the equilibrium is efficient, that is, $x^e = x^*$.

We shall now see what happens when F is positive, that is, hospital production takes place with a noncontractible input. Substituting the expression for r^e (20) into eq. (19), we get that

$$B_x(x^e) - 1 - \frac{F}{x^e}\left(1 - \frac{\beta}{1-\beta} B_x(x^e)\right) + \frac{F}{F + 2x^e} = 0. \tag{21}$$

Since we know that when $F = 0$, $x^e = x^*$ and, hence, $B_x(x^e) = 1$, we can conclude that:

$$\frac{dx^e}{dF} = -\left(\frac{\beta}{1-\beta} - \frac{1}{2}\right)\frac{1}{x^e B_{xx}(x^e)} \tag{22}$$

at $F = 0$.

We can therefore state the following:

PROPOSITION 2: *When F is small, $x^e \geq x^*$ ($x^e \leq x^*$) if $\beta \geq \frac{1}{3}$ ($\beta \leq \frac{1}{3}$).*

Proof. The proof follows directly from eq. (22) and from the fact that at $F = 0$, $x^e = x^*$. □

The more interesting case, however, is when F is not necessarily small. We can get more insight about the size of x^e relative to x^* by looking at dW_i/dr_i at $r_1 = r_2 = r^* \equiv (1 - 2\beta)/(1 - \beta)$. If $r = r^*$ then we know that $x_1 = x_2 = x^*$ and, hence, $B_x(x^*) = 1$. Substituting r^* and x^* into eq. (14), we get

$$\frac{dW_i}{dr_i} = \frac{(1-\beta)}{\beta B_{xx}(x^*)}\cdot\frac{F}{2x}\left(2\left(\frac{1-2\beta}{1-\beta}\right)\left(\frac{F}{2x^*} + 1\right) - 1\right). \tag{23}$$

If eq. (23) is greater than zero, then at r^*, both players would like to increase their r, and if eq. (23) is less than zero, both players would like to decrease their r. However, given that when r is increased (equally by both players), x would also increase, we can state the following proposition:

PROPOSITION 3: *Suppose that $\beta < \bar{\beta}$ and, hence, an interior equilibrium exists, then: $x^e \leq x^*$ if $\beta \leq (F + x^e)/(2F + 3x^e)$ and $x^e \geq x^*$ if $\beta \geq (F + x^e)/(2F + 3x^e)$, where x^e is given by eq. (21).*

Proof. The proof follows directly from the fact that if $\beta \leq (F + x^e)/(2F + 3x^e)$, then the left side of eq. (23) is negative, and if $\beta \geq (F + x^e)/(2F + 3x^e)$, the left side of eq. (23) is positive. □

Corollary. If $\beta < \frac{1}{3}$, then $x^e < x^*$.

Figure 2 summarizes what we know about the equilibrium x and r. As can be seen from Figure 2, the properties of equilibrium r and x are determined jointly by the size of β and the size of F. When F is small, then too little care will be provided in equilibrium, that is, $x^e < x^*$, if $\beta < \frac{1}{3}$, and too much care will be produced if $\beta > \frac{1}{3}$. However, when F gets large, then even with $\beta > \frac{1}{3}$ (but not greater than $\frac{1}{2}$), too little care will be provided relative to the socially optimal amount. Furthermore, when F is small, equilibrium contracts will have a retrospective payment only if $\beta < \frac{1}{2}$. However, when F gets large, also with $\beta > \frac{1}{2}$, equilibrium contracts will have a retrospective payment.

Interestingly, one effect of reliance on cost allocation is to increase the use of cost-related component in payment systems (r) in comparison with when per patient costs are contractible. In the case of all costs contractible, payers would set $r = 0$ when $\beta \geq \frac{1}{2}$. The resulting x^o would be the minimum x achievable within the bounds of the payment system parameters. Since $x^* < x^o$ when $\beta > \frac{1}{2}$, this is the best feasible allocation. We can conclude from this then that if $r_1^e = r_2^e = 0$ when cost allocation is used, this set of contracts cannot be altered to increase social welfare.

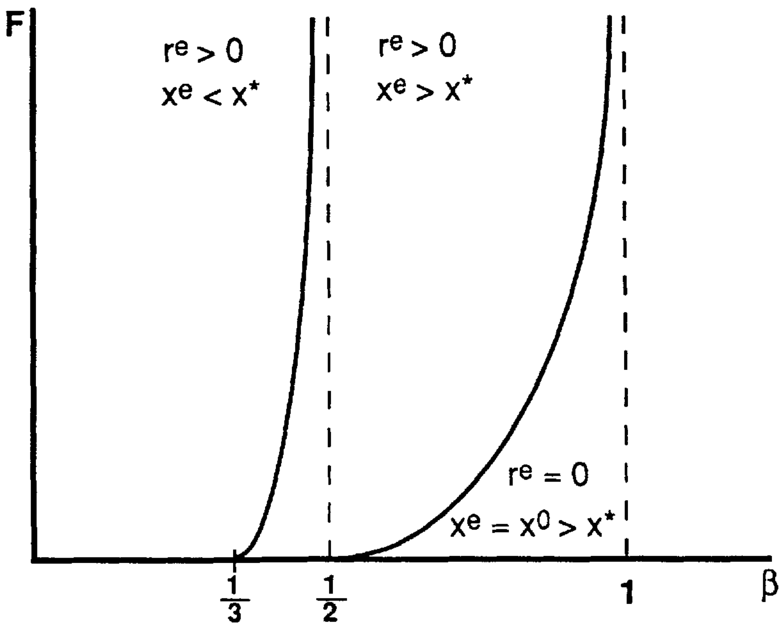

FIGURE 2. PROPERTIES OF EQUILIBRIUM.

6. Conclusion

Burner et al. (1992) estimate that $360 billion was spent on hospital care in 1993. Almost all of this will be governed by a contract between a third-party payer and the hospital. There are few examples of principal/agent relations that are more important than the payer/hospital relation in terms of the share of national income in question (about 5%) or in terms of the effect on consumer welfare.

The present direction of health reform in the United States is to encourage competition among payers by defining a standard insurance benefit, requiring choice of insurer and possibly eliminating tax subsidy for premium contributions above those for the standard coverage. The potential for fundamental reform of the health insurance function in the United States motivated us to study the properties of various payment systems in the presence of payer competition.

In this paper we show that competition between payers will not lead to efficient hospital production in the presence of noncontractible inputs. Payers will attempt to use cost-allocation rules to cost shift to their rivals. The resulting equilibrium will be characterized by payment systems that may be "too prospective," leading to underprovision of the variable input, or "not prospective enough," with overprovision of the variable input depending on the relative importance of the noncontractible input and the weight given to patient benefit in hospital input decisions.

The double-edged effect of cost shifting may surprise many people whose experience with cost-allocation rules and cost shifting dates from an era when payers were moving away from systems paying entirely on the basis of cost allocation. It is true that when your rival uses cost allocation only, you can cost shift by paying more prospectively and encouraging the hospital to allocate costs to your rival. But this is only part of the story with cost shifting. Even if your rival pays only with a prospective payment, you may cost shift by including a cost-allocation component in your payment, in effect using some of your rival's prospective payment to pay for services for your patients. The relative strength of these two forces will determine the direction of the inefficiency characterizing equilibrium.

Our analysis is based on a model in which two payers chose from among a family of payment systems referred to as "mixed" payment systems with a purely prospective and a purely cost-based system at the endpoints of a continuum. Although some public-sector contracts are mixed in this sense, private insurers are not observed to combine prospective payment with cost allocation. Why not? One explanation is that, within our model, parameter values for payers and hospitals are such that only prospective payment is a possible equilibrium.

We believe a better explanation rests on two points. First, our model is incomplete in many respects. To mention one factor, when hospitals compete for patients by setting treatment intensity, a payer may set payment to be more prospective to restrain this effect, tending to move the equilibrium payment contract more toward prospective payment. Furthermore, cost allocation, prospective payment, and their combinations are one family of feasible contracts in the presence of a noncontractible input. Paying prices for certain contractible inputs at a level sufficient to cover all costs represents another approach. Per diem payments that pay a hospital simply on the basis of a price per day is one simple example of such a contract. Future research should consider the character of equilibria when these contracts are permissible as well. Second, the private insurance market is changing so rapidly that anyone would be hard pressed to argue that it is in equilibrium. Both commercial insurance companies and Blue Cross plans are moving away from their different historical contract forms to new arrangements. In the new world of more uniform regulation of provider payment and contracting that might result from health reform, innovative forms of contract can be expected to appear.

The complexity of the incentives created by contracts in the presence of noncontractible inputs make it difficult to draw clear conclusions for public policy on the basis of a simple model. If a regulator knew the trade-off implicit in hospital decision making between patient welfare and profits, the right level of cost allocation component in a payment system could simply be dictated. National health policy is headed in the direction of prohibiting insurer competition over many dimensions of insurance coverage for patients; it is natural in light of our work here to consider whether competition in the form of payment to providers should be regulated as well.

References

Berman, H.J. and L.E. Weeks, 1982, *The Financial Management of Hospitals*, 5th ed., Chicago: Health Administration Press.

Bernheim, D.B. and M.D. Whinston, 1986, "Common Agency," *Econometrica*, 54, 923–944.

Burner, S.T., D.R. Waldo, and D.R. McKusick, 1992, "National Health Expenditures Projections Through 2030," *Health Care Financing Review*, 14, 1–29.

Clinton, W., 1993, "Health Security Act."

Congressional Research Service, January 1993, *Medicaid Source Book: Background Data and Analysis (1993 Update)*, Washington, DC.

Dranove, D., 1988, "Pricing by Non-Profit Institutions: The Case of Hospital Cost-Sharing," *Journal of Health Economics*, 7, 47–57.

Ellis, R.P. and T.G. McGuire, 1986, "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, 5, 129–151.

——— and ———, 1993, "Supply-Side and Demand-Side Cost Sharing in Health Care," *Journal of Economic Perspectives*, 7, 135–151.

Enthoven, A. and R. Kronick, 1991, "Universal Health Insurance Through Incentives Reform," *Journal of the American Medical Association*, 265, 1532–1537.

Foster, R., 1985, "Cost-Shifting Under Prospective Reimbursement and Prospective Payment," *Journal of Health Economics*, 4, 261–275.

Frank, R.G. and J.R. Lave, 1989, "A Comparison of Hospital Responses to Reimbursement Policies for Medicaid Psychiatric Patients," *Rand Journal of Economics*, 20, 588–600.

——— and D. Salkever, 1991, "The Supply of Charity Services by Non-Profit Hospitals: Motives and Market Structure," *Rand Journal of Economics*, 22, 430–445.

Frankfurt, D., 1993, "The Medicare DRGs: Efficiency and Organizational Rationality," *Yale Journal of Regulation*, 10, 273–346.

Glaser, W.A., 1987, *Paying the Hospital*, San Francisco: Jossey-Bass Publishers.

Goodall, C., 1990, "A Simple Objective Method for Determining a Percent Standard in Mixed Reimbursement Systems," *Journal of Health Economics*, 9, 253–271.

Hay, J.W., 1983, "The Impacts of Public Health Care Financing Policies on Private Sector Hospital Costs," *Journal of Health Policy, Politics and Law*, 7, 945–952.

Hoy, E., R.E. Curtis, and T. Rice, 1991, "Change and Growth in Managed Care," *Health Affairs*, 10, 18–36.

Loeb, M. and W.A. Magat, 1979, "A Decentralized Method for Utility Regulation," *The Journal of Law and Economics*, 22, 399–404.

Ma, C.A., 1994, "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics & Management Strategy*, 3, this issue.

——— and T.G. McGuire, "Paying for Joint Costs in Health Care," *Journal of Economics & Management Strategy*, 2, 71–96.

Pope, G.C., 1989, "Hospital NonPrice Competition and Medicare Reimbursement Policy," *Journal of Health Economics*, 8, 147–172.

Rogerson, W., 1994, "Choice of Treatment Intensities by a Nonprofit Hospital Under Prospective Pricing," *Journal of Economics & Management Strategy*, 3, 7–51.

Sloan, F. and E. Becker, 1984, "Cross Subsidies and Payment for Hospital Care," *Journal of Health Politics, Policy and Law*, 8, 660–683.