Cost and Quality Incentives in Health Care: Altruistic Providers

Ching-to Albert Ma

Department of Economics
Boston University
270 Bay State Road
Boston
Massachusetts 02215
ma@bu.edu

December, 1997

JEL Classification Numbers: I10, I18, L10, L50
Keywords: Cost, Quality, Dumping, Prospective Payment, Cost Reimbursement, Hospital, Health Care

Abstract

This paper compares the cost and quality incentive effects of cost reimbursement and prospective payment systems in the health industry when providers are altruistic. Providers' behavioral rule is governed by a desire to maximize a weighted sum of profit and consumers' health benefit. When providers exert costly efforts to enhance quality and reduce cost, the first best—a regime in which efforts are contractible—can be implemented by prospective payment. Cost reimbursement does not implement the first-best efforts. Necessary conditions are derived for the implementation of the first best by a combination of prospective and cost reimbursement payments when the provider can refuse services to expensive patients.

1 Introduction

In the health care market, consumers often pay very little of the expenses when they consume medical care. Many consumers are insured by their employers, the government, or have purshased private health insurance contracts. These insurance policies insulate the consumers from the risks of medical expenditure when they become ill. The consequence of insurance on consumers' behavior—moral hazard—is well recognized, and the economics literature has studied extensively the various kinds of inefficiency that may result. Studies that focus on the demand side of the health market often ignore or make simplifying assumptions on health care providers' responses. Clearly, the outcome in any market is generated by the interaction of both demand and supply forces. Although consumes often do not directly pay providers at the time of service, providers do receive payments from the insurers. A good understanding of the health market therefore relies as much on the analysis of the way providers are compensated as on the way consumers (and physicians) decide on medical services.

Payment systems—the method insurers use to pay providers—occupy a central position in the supply side of the health industry. Indeed, popular wisdom often blames the exorbitant growth in the share of health care expenditures in the gross national product in the United States on the federal government's accommodating cost reimbursement policies in the past few decades. The change in Medicare's payment policies almost fifteen years ago was an attempt to avoid the supposedly perverse incentives under cost reimbursement policies. When Medicare introduced the prospective payment system, which pays a fixed "price" per discharge according to the patient's diagnosis (Diagnostic Related Group or DRG), it was hoped that inefficiency in the supply side will be minimized, since providers then would be made responsible for the production costs.

The effect of any given payment system depends critically on a provider's reactions.

Such reactions in turn are the results of a provider's attempt to optimize on an objective function within a set of feasible actions. In many industries, firms are expected to maximize profits, since the majority of them are public companies. In the health industry in the U.S. as well as many countries, the majority of providers are nonprofit organizations. In fact, Arrow (1963) more than thirty years ago pointed out that health care providers may be interested in their patients' welfare, and behave in an altruistic way. The same idea has also been emphasized in Eisenberg (1986). A payment system that relies on the reactions of a pure profit-maximizing provider may lead to unintended consequences when the provider's objective includes consumer benefits. Second, and perhaps more importantly, a provider may be motivated to explore new ways to react to a particular payment system. Again, the effect of a payment system may critically depend on whether a provider may or may not perform a certain kind of activities.

In this paper, I assume that the provider's preferences depend on its profits as well as consumers' benefits. I also allow the provider to refuse supplying services to consumers; if a patient is not benefiting the provider, in the sense of raising its utility, services to the patient may be withheld. The contribution of this paper is to draw attention to the different *incentive tradeoffs* between two payment methods, namely cost reimbursement and prospective payment. I model the *interaction* of cost and quality incentives when the provider is partially altruistic and when a counter-strategy of *dumping* patients is available to the regulated health care provider. Given a payment system, the agent determines the mix of efforts on cost reduction and quality enhancement activities to maximize its utility.

In the first version of our model, the regulated health care provider can only allocate its efforts between cost reduction and quality enhancement; dumping is ruled out. While cost reimbursement provides no cost-reduction incentive, quality enhancement incentives will still be present. A "constrained" optimal level of quality, in the sense that cost-reduction activitiy is set at the minimum level, can be implemented by cost reimbursement. Prospective payment does achieve the efficient allocation of efforts between cost and quality. In an earlier paper (Ma 1994), I proved this result when the provider maximizes only its profit. In this paper the provider values also consumer benefit, which is a function of care quality; hence, the earlier result cannot be applied directly. In fact, when the provider is altruistic, the level of prospective payment must be set to "neutralize" the effect of altruism, so that the provider does not value excessively the benefit of quality. I also show that as the degree of altruism increases, the efficient prospective payment resembles an average-cost pricing rule. The analysis is also extended to consider multiple services.

The above result serves as a benchmark for the evaluation of the prospective and cost reimbursement policy regimes when the provider can refuse patients. Because of altruism, the provider does have a tendency to supply services: patient benefits are valued by the provider, and they are only available when the provider treat the patients. Nevertheless, when costs become too high, the disutility from supplying service may still outweigh the benefit, and the provider may dump the patient. Because the hospital never refuses a patient under cost reimbursement, the policy can again guarantee a constrained efficient level of quality. Under prospective payment, the hospital will refuse very expensive patients. The comparison between the welfare levels achieved by prospective payment and cost reimbursement is ambiguous. I then explore a "mixed" reimbursement method, a system in which low-cost patients are paid by a fixed price, but high-cost patients by cost reimbursement. I derive necessary conditions for the implementation of the efficient cost and quality efforts.

A number of earlier papers have discussed the limitations and incentive problems of the prospective payment system. Newhouse (1983, 1984), Pauly (1984), and Russell (1989) contain both theoretical and empirical assessments. Ellis and McGuire (1986) question whether the prospective payment system alone can adequately align differences in preferences between doctors and patients; they show that a combination of prospective payment and (partial) cost reimbursement will be necessary when the doctor is an imperfect agent for the patient.

This paper is similar in this the provider's objective is not pure profit maximization. Dranove (1987) analyzes the dumping and specialization incentives under prospective payment, while Allen and Gertler (1991) demonstrate that when patient heterogeneity exists within a DRG, some type of patients must receive inefficient quality under prospective payment. This paper builds on my work in 1994, where I have analyzed cost and quality incentives of payment systems when the provider seeks to maximize profits.[1]

After the introduction of the prospective payment system fifteen years ago, economists already pointed out some of its inadequacies, but few actually analyzed the balance between cost reduction and quality enhancement efforts allocations. Nevertheless, Shleifer (1985), using a model of yardstick competition, demonstrates that prospective payments can induce the efficient cost decisions by hospital. Nevertheless, the Shleifer paper did not include quality decisions. Pope (1989) studies an imperfectly competitive market in which hospitals use quality to compete for patients; in that paper, the hospital can enhance quality and reduce "managerial slack." While Pope shows that competition may reduce managerial slack but result in excessive quality, the consequences of provider altruism and dumping are not considered. I do not consider competition, but the quality and cost efforts in this paper are similar to the hospital's available strategies in Pope (1989). Recently Ellis (1998) studies a duopoly model in which hospitals can employ dumping as well as care intensity (quality) to compete for and discriminate against patients. Although dumping, creaming, and skimping are considered, Ellis (1998) does not allow cost reduction efforts.

The rest of the paper is organized as follows. The model is described in section 2; an efficiency benchmark is derived there. In the following section, I show that when cost reduction and quality enhancement are the only activities for the hospital, prospective payment implements the efficient allocation even when the provider is partially altruistic. In Section 4,

[1]See also related papers by Chalkley and Malcomson (1995) and (1998).

I consider cost reimbursement, and then in the next section, the analysis is extended to the case of multiple services. Section 6 studies dumping and provider altruism. I show that a necessary and sufficient condition for the implementation of the first best by prospective payment is the same whether the provider is altruistic or not. Then I introduce a mixed payment system and derive a set of necessary conditions for the implementation of the first best by the mixed system. The last section contains some concluding remarks.

2 A Model of an Altruistic Provider

I set up a model of a partially altruistic, regulated health care provider. The analysis concentrates on the supply side; I assume that consumers are fully insured, and ignore the possible incentive effects of demand-side cost sharing. Because consumers are assumed to be fully insured, they base their choice of a health care provider on the provider's quality of care. Accordingly, consumer demands at the provider are an increasing function of its care quality.

A provider can use its "effort" to improve quality and to reduce its marginal cost of services. Efforts, however, are unobservable to a payer, and hence the provider cannot be directly compensated for its choice of efforts. The variables q and e denote efforts a provider can direct to quality enhancement and cost reduction dimensions respectively; minimum effort levels are normalized at zero. These efforts impose a total disutility of $\gamma(q+e)$ to the provider. The function γ is increasing and convex.

A higher care quality requires a higher effort input by the provider. By a slight abuse of notation, I also use q to denote the provider's quality of care at the provider. The provider's unit cost of treating a patient is $c(e)$ when the cost reduction effort is e. It is assumed that $c(e)$ is decreasing and convex. Later, in Section 6, I will allow the cost to be random; in this case, a higher cost-reduction effort will reduce the expected cost. Also, let the increasing and concave function $\mu(q)$ be the provider's demand when the quality of services is q.

Two reimbursement methods are the main focus of the paper. Under cost reimbursement, the provider's actual production costs are completely paid for; in addition, the provider may be paid a nonnegative margin, m, so that if the provider's cost of treating a patient is c, it receives a total reimbursement of $c+m$. The margin may be used to motivate the provider to supply effort on quality enhancement. Under prospective payment, the provider receives a fixed payment per discharge, p, independent of the total costs of treatment. The provider's net profit under cost reimbursement and prospective payment are respectively:

$$(c(e)+m)\mu(q) - c(e)\mu(q) - \gamma(q+e) = m\mu(q) - \gamma(q+e) \tag{1}$$

$$p\mu(q) - c(e)\mu(q) - \gamma(q+e). \tag{2}$$

Although in the paper, the provider's utility depends on profit as well as patients' benefits (to be defined shortly), it must still earn a nonnegative profit. So nonnegative profit constraints will be used throughout the paper.

The gross benefit produced by the provider generally depends on the number of consumers using the provider and the quality of care these clients receive. But since the provider's demand is an increasing function of its quality, I simply write the gross benefit as a function of quality alone, $W(q)$. The function W is increasing and concave. I use a general gross benefit function. In many applications, benefits are taken to be consumer surplus. The function W includes that as a special case, but is sufficiently general to incorporate other policy issues that may be important in the health industry. For instance, because of risk sharing or other institutional and informational distortions, the appropriate benefit measure may be bigger than consumer surplus.

A regulator or a payer uses a payment policy to reimburse the provider for its costs of treating patients. This regulator or payer may be a public agency or a private insurance company. In any case, the regulator's preferences are given by the difference between consumer benefit and the total cost of production:

$$W(q) - \mu(q)c(e) - \gamma(q + e). \tag{3}$$

If $W(q)$ is taken to represent gross consumer surplus, then (3) measures the sum of net consumer surplus and producer surplus.[2] The provider's preferences are given by the sum of profits and $\theta W(q)$, where $0 < \theta < 1$ measures the degree of altruism.

As a benchmark, the efforts that maximize the regulator's objective function (3) are called the efficient allocation of efforts. Denote the first best by (q^*, e^*). I assume that the first-rder conditions are sufficient to characterize the first best:

$$W'(q^*) - \mu'(q^*)c(e^*) - \gamma'(q^* + e^*) = 0 \tag{4}$$

$$-\mu(q^*)c'(e^*) - \gamma'(q^* + e^*) = 0. \tag{5}$$

3 Prospective Payment

In this and the next two sections, I assume that the provider cannot refuse providing treatments to consumers. Therefore, for any given payment system, the provider may only pick the quality and cost efforts to maximize its utility. In this section, I study the efficiency properties of prospective payment.

The provider's objective function under prospective payment is given by

$$(p - c(e))\mu(q) - \gamma(q + e) + \theta W(q).$$

[2]The definition in (3) ignores distribution issues and concentrates on efficiency. See Ma (1994) for a discussion when the regulator's preferences are written as a weighted sum of consumer benefits and profits.

For a given prospective payment p, the provider chooses q and e to maximize

$$(p - c(e))\mu(q) - \gamma(q + e) + \theta W(q) \tag{6}$$

subject to

$$(p - c(e))\mu(q) - \gamma(q + e) \geq 0. \tag{7}$$

The constraint ensures that the provider makes a nonnegative profit. Observe that the objective function is increasing in p.

I now consider the implementation of the first best, (q^*, e^*), by a prospective payment system. In other words, does there exist p so that q^* and e^* will be optimally chosen by the provider? I begin by considering a relaxed problem: the unconstrained maximization of (6). Let $q(p)$ and $e(p)$ be the solution, and $F(p)$ the maximized value of (6) without the constraint. Clearly, by the Envelope Theorem, $F'(p) = \mu(q(p)) > 0$.

If the constraint (7) does not bind, then the first best can be implemented in a straightforward way. First, for a given p the first-order conditions for the maximization of (6) are:

$$(p - c(e))\mu'(q) - \gamma'(q + e) + \theta W'(q) = 0 \tag{8}$$

$$-c'(e)\mu(q) - \gamma'(q + e) = 0. \tag{9}$$

Then to implement the first best (q^*, e^*), simply set p to

$$p^* \equiv \frac{(1 - \theta)W'(q^*)}{\mu'(q^*)}. \tag{10}$$

When p is set to p^*, then (q^*, e^*) satisfies the first-order conditions (8) and (9), and the first best is implemented.

Since the provider completely internalizes costs, it has an incentive to lower them efficiently. Nevertheless, the objective function includes not only profit but also the social benefit of quality. The prospective payment to implement any pair of quality and cost reduction efforts must take this into account. Therefore, the prospective payment for the implementation of the first best, p^*, is a function of θ, which measures the provider's altruism. Indeed, because the provider values quality due to altruism, the price to implement the efforts, p^*, is decreasing in the degree of altruism, θ.

Given that $p = p^*$, the equilibrium efforts are q^* and e^* if the provider is able to make nonnegative profit (so that (7) does not bind). Under what circumstances will the implementation of the first best simply by p^* be feasible? Consider the profit expression

$$(p^*(\theta) - c(e^*))\mu(q^*) - \gamma(q^* + e^*), \tag{11}$$

where the fact that p^* is a function of θ is now emphasized. By construction, when θ changes, the efforts will remain at the first best, since $p^*(\theta)$ implements the first best. Hence, this profit function is decreasing in θ. When θ approaches 1, $p^*(\theta)$ tends to 0 and the value of (11) must become negative. Therefore, it must be concluded that the implementation of the first best by p^* is feasible if and only if θ is sufficiently low. More precisely, define $\overline{\theta}$ by $(p^*(\overline{\theta}) - c(e^*))\mu'(q^*) - \gamma'(q^* + e^*) = 0$. Then for $\theta > \overline{\theta}$, the nonnegative profit constraint becomes binding.

I next consider the case when $\theta > \overline{\theta}$. When $p < p^*(\theta)$ is used, the provider cannot break even. Therefore, given that $p < p^*(\theta)$, the maximization of (6) when (7) binds must be considered. The characterization of the solution to this maximization program is as follows. Since (7) binds, the program can be rewritten as the maximization of $\theta W(q)$ subject to (7) holding as equality. As $\theta > 0$ and W is increasing, this is the same as the maximization of q subject to $(p - c(e))\mu(q) - \gamma(q + e) = 0$. The first-order conditions are given by

$$1 + \lambda[(p - c(e))\mu'(q) - \gamma'(q + e)] = 0$$

$$-c'(e)\mu(q) - \gamma'(q + e) = 0$$

$$(p - c(e))\mu(q) - \gamma(q + e) = 0,$$

where λ is the multiplier. Clearly, the solution $(q(p), e(p))$ is given by the last two equations since λ can be (freely) chosen to satisfy the first equation. The solution is easily described. Because the objective is to maximize q subject to the zero-profit constraint, the "cost" $c(e)\mu(q) + \gamma(q + e)$ must be minimized, and any excess profit spent on expanding q. It is important to note that the solution is independent of the value of θ (as long as $\theta > \overline{\theta}$ so that $p < p^*(\theta)$ and (7) binds).

In summary, when the value of θ is higher than $\overline{\theta}$, then the social welfare function (3) must be maximized subject to

$$-c'(e)\mu(q) - \gamma'(q + e) = 0 \tag{12}$$

$$(p - c(e))\mu(q) - \gamma(q + e) = 0. \tag{13}$$

I now claim that these two constraints are not binding. Consider again the first best (q^*, e^*), which maximizes the social welfare function (3). Clearly, (q^*, e^*) satisfies (12). The value of p does not affect (3), and hence it can be set at $p^*(\overline{\theta})$. Again, by the definition of $\overline{\theta}$, (q^*, e^*) satisfies (13) when p is set at $p^*(\overline{\theta})$.

Proposition 1 *When the provider's degree of altruism is small ($\theta \leq \overline{\theta}$), the first-best efforts can be implemented by the prospective payment $(1-\theta)W'(q^*)/\mu'(q^*)$. In this case, the provider makes strictly positive profits. When the provider's degree of altruism is large ($\theta > \overline{\theta}$), the first best can be implemented by the prospective payment $(1-\overline{\theta})W'(q^*)/\mu'(q^*)$. In this case, the provider makes zero profit, and the prospective payment is identical to average-cost pricing, and independent of the degree of altruism as long as $\theta > \overline{\theta}$.*

4 Cost Reimbursement

I now study the efficiency properties of cost reimbursement. Since the cost of treatment will be paid for, the provider's objective function can be written as

$$m\mu(q) - \gamma(q+e) + \theta W(q).$$

Hence, the provider chooses quality and cost efforts, q and e to maximize

$$m\mu(q) - \gamma(q+e) + \theta W(q) \tag{14}$$

subject to

$$m\mu(q) - \gamma(q+e) \geq 0. \tag{15}$$

When costs are reimbursed completely, the provider will not have any incentive to put in any cost-reduction effort. So $e = 0$. But the margin m allows the provider to make a profit from providing treatments. The profit margin therefore is an incentive for the provider to increase its demand by supplying care at a higher quality. How can the margin be set appropriately?

I establish now another benchmark. The "constrained" social optimum, $q^\dagger$, is the optimal level of quality when e is set at its minimum level. Formally, $q^\dagger$ solves

$$\max_q W(q) - c(0)\mu(q) - \gamma(q)$$

and is given by the first-order condition

$$W'(q^\dagger) - c(0)\mu'(q^\dagger) - \gamma'(q^\dagger). \tag{16}$$

Suppose that the nonnegative profit constraint (15) does not bind. Then the constrained optimal quality level can be implemented by setting m to $m(\theta)$, where

$$m(\theta) \equiv -c(0) + \frac{(1-\theta)W'(q^\dagger)}{\mu'(q^\dagger)}. \tag{17}$$

Given $m(\theta)$, a provider's choice of quality to maximize (14) is $q^\dagger$. The margin $m(\theta)$, which implements $q^\dagger$, is decreasing in θ. Moreover, given that q is set at $q^\dagger$, the provider's profit $m(\theta)\mu(q^\dagger)-\gamma(q^\dagger)$ is decreasing in θ. Hence, when θ is sufficiently high, the nonnegative profit constraint must become binding. This critical value, $\hat{\theta}$, is given by $m(\hat{\theta})\mu(q^\dagger) - \gamma(q^\dagger) = 0$; that is, for $\theta > \hat{\theta}$, the nonnegative profit constraint (15) must bind. Conversely, for $\theta < \hat{\theta}$, the provider makes a positiive profit, and the constrained optimal quality can be implemented by $m(\theta)$ defined in (17) above.

Suppose now that $\theta > \hat{\theta}$. If any $m < \hat{\theta}$ is used, the provider will not break even. Can the constrained optimal quality $q^\dagger$ still be implemented? To answer this question, consider the maximization of (14) subject to the binding constraint (15). By substituting the constraint into the objective function, one obtains the following equivalent program: the maximization of $\theta W(q)$ subject to (15) holding as an equality. The solution is given simply by $m\mu(q) - \gamma(q) = 0$, which therefore defines the level of quality that will be implemented for a given m when $\theta > \hat{\theta}$. Notice that this implementation condition is independent of the value of θ as long as it is higher than $\hat{\theta}$. It follows that the optimal choice of the margin m for $\theta > \hat{\theta}$ must be $m(\hat{\theta})$. When this value is used for m, the constrained optimal quality level is implemented, and the provider just breaks even.

Proposition 2 *When the provider's degree of altruism is small ($\theta \leq \hat{\theta}$), the constrained optimal quality effort can be implemented by the cost margin* $(1-\theta)W'(q^\dagger)/\mu'(q^\dagger) - c(0)$. *In this case, the provider makes strictly positive profits. When the provider's degree of altruism is large ($\theta > \hat{\theta}$), the constrained optimal quality effort can be implemented by the cost margin* $(1-\hat{\theta})W'(q^\dagger)/\mu'(q^\dagger) - c(0)$. *In this case, the provider makes zero profit, and the margin over cost is just sufficient to cover the average cost of effort and independent of the degree of altruism.*

5 Multiple Services and Payments

In the previous two sections, I have characterized the efficiency properties of the prospective payment and cost reimbursement systems. The provider's partial altruism mainly affects the provider's profit level when quality and cost-reduction efforts are implemented. I now extend the analysis to consider the case of two services, and the index $i = 1, 2$ denotes them. Let $\mu_i(q_i)$ denote the demand function for service i when quality of service i is q_i, $i = 1, 2$. Similarly, let the social benefit from service i be $W_i(q_i)$. Let the marginal cost of service i be $c_i(e_i)$, where e_i is the cost-reduction effort for service i. Finally, the total disutility for the quality and cost-reduction efforts for the two services is given by $\gamma(q_1 + q_2 + e_1 + e_2)$.

The first best is defined by q_i^* and e_i^*, $i = 1, 2$, that maximize the social welfare function

$$W_1(q_1) + W_2(q_2) - \mu_1(q_1)c_1(e_1) - \mu_2(q_2)c_2(e_2) - \gamma(q_1 + q_2 + e_1 + e_2). \tag{18}$$

The constrained optimal quality levels, $q_1^\dagger$ and $q_2^\dagger$, are those that maximize (18) when e_1 and e_2 are set at zero.

Under prospective payment, the provider's objective function is

$$\sum_{i=1}^{2}\{[p_i - c_i(e_i)]\mu_i(q_i) + \theta W_i(q_i)\} - \gamma(q_1 + q_2 + e_1 + e_2), \tag{19}$$

where p_1 and p_2 are the prospective prices for serivces 1 and 2 respectively. Under cost reimbursement, the provider's objective function is

$$\sum_{i=1}^{2}\{m_i\mu_i(q_i) + \theta W_i(q_i)\} - \gamma(q_1 + q_2 + e_1 + e_2), \tag{20}$$

where m_1 and m_2 are the margins above costs for serivces 1 and 2 respectively. In each case, the provider faces the nonnegative profit constraints:

$$[p_1 - c_1(e_1)]\mu_1(q_1) + [p_2 - c_2(e_2)]\mu_2(q_2) - \gamma(q_1 + q_2 + e_1 + e_2) \geq 0$$

$$m_1\mu_1(q_1) + m_2\mu_2(q_2) - \gamma(q_1 + q_2 + e_1 + e_2) \geq 0.$$

If the nonnegative profit constraints do not bind, the first-best quality and cost-reduction efforts can be implemented in much the same way as the case of a single service. That is, simply set the prospective prices p_1^* and p_2^* to

$$p_i^* \equiv \frac{(1-\theta)W_i'(q_i^*)}{\mu_1'(q_1^*)}, \quad i = 1, 2. \tag{21}$$

Under these prices, and when the efforts are q_i^* and e_i^*, the provider's profit function becomes

$$\sum_{i=1}^{2} \left[\frac{(1-\theta)W_i'(q_i^*)}{\mu_i'(q_i^*)} - c_i(e_i^*) \right] \mu_i(q_i^*) - \gamma(q_1^* + q_2^* + e_1^* + e_2^*). \tag{22}$$

This expression is again decreasing in θ, as in the case of a single service. So for sufficiently high values of θ, the expression in (22) becomes negative and the nonnegative profit constraint binds; the threshold level, $\hat{\theta}$, is given by that level of θ so that (22) is equal to zero. Let $p_i^*(\hat{\theta})$, $i = 1, 2$, denote the corresponding prices; that is, expressions in (21) when θ is set at $\hat{\theta}$. As in the case of a single service, for those values of θ higher than $\hat{\theta}$, the cost-reduction efforts continue to be efficient, and the efficient qualities can be implemented by setting the prices to $p_i^*(\hat{\theta})$.

The analysis for the implementation of the constrained optimal quality effort under cost reimbursement is also straightforward, and I will only outline the results. For values of θ that are sufficiently small, the margins that implement the constrained optimal quailities are:

$$-c_i(0) + \frac{(1-\theta)W_i'(q_i^\dagger)}{\mu'(q_i^\dagger)}.$$

For some critical value of θ, the provider's profit will become zero if the above margins are used. For those values that are higher than the critical value, then use the margins that implement the constrained optimal qualities for the critical value; these margins together with the actual cost reimbursements will pay for provider's average cost.

6 Dumping and Altruism

I now consider a more complex environment. In this section, I allow the provider to choose which patients to serve after observing their treatment costs. In this version I will assume that the costs to treat patients vary, and are described by a random variable. Formally, cost variation for each patient is described by the distribution function $F(c; e)$ with density function $f(c; e)$ on the support $[0, \hat{c}]$; patients' costs are distributed independently. The distribution and density functions depend on the provider's cost effort. By a slight abuse of notation, I use $c(e)$ to denote the expected value of c when cost reduction effort is e. Then the first best is again given by (4) and (5).

Refusing to treat patients is certainly unattractive when the provider is paid by cost reimbursement; treating a patient earns the margin m. Therefore, results on the efficiency properties of the cost reimbursement system derived in the previous sections continue to apply. The analysis in this section is on the prospective payment system.

Suppose that the prospective price is p. I now determine a "threshold" cost value, x: when a patient's treatment cost is higher than x, the provider refuses to supply treatment. I begin by writing write down the provider's payoff:

$$\mu(q) \int_0^x (p - c) dF(c; e) - \gamma(q + e) + \theta W(q) F(x; e),$$

which can be rewritten as

$$\mu(q) \int_0^x \left[p + \theta \frac{W(q)}{\mu(q)} - c \right] dF(c; e) - \gamma(q + e).$$

Here, a quality of q attracts a total of $\mu(q)$ consumers. The provider is also altruistic and values the social benefit $W(q)$. The social benefit per consumer is $\theta W(q)/\mu(q)$. I assume that if the provider agrees to treat a patient, then under prospective payment, the gain in utility is $p + \theta W(q)/\mu(q)$. When a provider is able to refuse treatment to a consumer, then a patient with a treatment cost higher than $p + \theta W(q)/\mu(q)$ will not be given service. In summary, the cost threshold for accepting a patient is given by $p + \theta W(q)/\mu(q)$.

Under prospective payment, when dumping is possible, a provider will choose q and e to maximize

$$\mu(q) \int_0^{p+\theta\frac{W(q)}{\mu(q)}} \left[p + \theta\frac{W(q)}{\mu(q)} - c\right] dF(c;e) - \gamma(q+e). \tag{23}$$

The first-order conditions with respect to e and q are, respectively,

$$\mu(q) \int_0^{p+\theta\frac{W(q)}{\mu(q)}} \left[p + \theta\frac{W(q)}{\mu(q)} - c\right] dF_e(c;e) - \gamma'(q+e) = 0 \tag{24}$$

$$\int_0^{p+\theta\frac{W(q)}{\mu(q)}} \left\{\mu'(q)\left[p + \theta\frac{W(q)}{\mu(q)} - c\right] + \mu(q)\theta\frac{d}{dq}\frac{W(q)}{\mu(q)}\right\} dF(c;e) - \gamma'(q+e) = 0. \tag{25}$$

The last condition simplifies to

$$\mu'(q) \int_0^{p+\theta\frac{W(q)}{\mu(q)}} \left[p + \theta\frac{W'(q)}{\mu'(q)} - c\right] dF(c;e) - \gamma'(q+e) = 0.$$

When dumping is possible, the provider generally does not internalize all production costs, lacking the full social incentive to reduce them. Nevertheless, given that the provider is altruistic, one may expect that the problem of dumping may not be too severe. The provider values consumer benefit, which is only available if the consumer receives treatment. Indeed, as the derivation of the treatment cost refusal threshold shows, for any given level of prospective payment, the consumer's cost will have to be strictly higher than the level of prospective payment for dumping to occur. Will this imply that the implementation of the first best is more likely than if the provider is not altruistic (the case of $\theta = 0$)?

If the provider is not altruistic, the implementation of the first best by prospective payment is feasible if and only if $W(q^*)/\mu(q^*) \geq \hat{c}$. This inequality says that the prospective payment to implement the first best when there is no dumping is actually higher than the maximum potential cost. Obviously, when this payment level is used, the provider will always supply treatment.

It turns out that this sufficient condition for the implementation of the first best is exactly the same when the provider is altruistic. Suppose the first best is implemented by some prospective payment. Then to avoid dumping, the prospective payment p must satisfy

$$p + \theta \frac{W(q^*)}{\mu(q^*)} \geq \hat{c}. \tag{26}$$

Furthermore, the first-order conditions for the provider's optimal choice of q and e, (24) and (25), must yield the same solution as those for the first best, (4) and (5). This implies that

$$p = (1 - \theta) \frac{W(q^*)}{\mu(q^*)}$$

But this result and inequality (26) mean that $W(q^*)/\mu(q^*) \geq \hat{c}$. So the condition for the implementation of the first best by prospective payment under dumping is the same whether the provider is altruistic or not. The intuition is actually simple. Although an altruistic provider may treat consumers even when their costs are higher than the prospective price, the altruistic motive must be completely realigned with the social incentive when the first best is implemented. As in the case when dumping is infeasible, the prospective price to implement the first best is a decreasing function of the degree of altruism θ. Because of this lower prospective price, the provider's treatment threshold, when the first best is implemented, is exactly the same whether dumping is possible or not, and this has to be higher than the upper support of the cost distribution.

Next, suppose that pure prospective payment will lead to the possibility of patient dumping (so that $W(q^*)/\mu(q^*) \geq \hat{c}$ is violated). I now explore the implementation of efficient efforts by a combination of prospective payment and cost reimbursement. The *mixed* payment system is defined by a triple, $(p, c^\dagger, m)$: if the provider incurs a cost c below $c^\dagger$, it will be paid p; above $c^\dagger$, $c + m$, where $m \geq 0$. Thus, for cost realizations below $c^\dagger$, the provider receives a fixed price p; above $c^\dagger$, a constant margin m over cost.

I have shown above that under a prospective payment of p, the cost threshold for accepting a patient is given by $p + \theta W(q)/\mu(q)$ when the provider chooses quality q. If a mixed system does not lead to dumping, then

$$c^{\dagger} \leq p + \theta \frac{W(q)}{\mu(q)}. \tag{27}$$

Under the mixed system and when dumping is possible, the provider's payoff is given by

(28)
$$\mu(q)\left\{\int_0^{\min[c^{\dagger},p+\theta\frac{W(q)}{\mu(q)}]}\left[p+\theta\frac{W(q)}{\mu(q)}-c\right]dF(c;e)+\int_{c^{\dagger}}^{\hat{c}}\left[m+\theta\frac{W(q)}{\mu(q)}\right]dF(c;e)\right\}-\gamma(q+e).$$

By construction, dumping does not occur when costs are higher than $c^{\dagger}$ because the provider is then paid by cost reimbursement, and receives a margin m per patient. This accounts for the second integral in the above expression. If costs are lower than $c^{\dagger}$, then the utility from treating the patient is $p + \theta W(q)/\mu(q)$. But dumping may actually occur if $c^{\dagger} > p + \theta W(q)/\mu(q)$. The highest level of cost for which dumping is not used and for which the provider is paid by prospective payment is therefore $\min[c^{\dagger}, p + \theta \frac{W(q)}{\mu(q)}]$. This explains the first integral of the above payoff.

I now examine the necessary conditions for the implementation of the first-best efforts e^* and q^* by the mixed payment system when dumping by the provider is possible. Suppose there is a mixed system $(p, c^{\dagger}, m)$ that implements (e^*, q^*). First, it must be the case that dumping does not occur:

$$c^{\dagger} \leq p + \theta \frac{W(q^*)}{\mu(q^*)}.$$

Therefore, it $\min[c^{\dagger}, p + \theta \frac{W(q)}{\mu(q)}] = c^{\dagger}$. Therefore, the first-order (right-hand) derivatives of (28) with respect to q and e can be obtained and evaluated under the assumption that the first-best efforts are chosen. When dumping does not occur, the profit function can be rewritten as

$$\mu(q)\left\{pF(c^{\dagger};e)+m[1-F(c^{\dagger};e)]-\int_0^{c^{\dagger}}c\,dF(c;e)\right\}-\gamma(q+e)+\theta W(q),$$

whose first-order derivatives with respect to q and e are:

$$\mu'(q)\left\{pF(c^\dagger;e)+m[1-F(c^\dagger;e)]-\int_0^{c^\dagger} cdF(c;e)\right\}-\gamma'(q+e)+\theta W'(q)$$

$$\mu(q)\left\{pF_2(c^\dagger;e)-mF_2(c^\dagger;e)-\int_0^{c^\dagger} cdF_2(c;e)\right\}-\gamma'(q+e).$$

If the first best (q^*, e^*) is implemented by some mixed system, then (27) is satisfied and the first best is given by the first-order condition:

$$\mu'(q^*)\left\{pF(c^\dagger;e^*)+m[1-F(c^\dagger;e^*)]-\int_0^{c^\dagger} cdF(c;e^*)\right\}+\theta W'(q^*)=\gamma'(q^*+e^*) \tag{29}$$

$$\mu(q^*)\left\{pF_2(c^\dagger;e^*)-mF_2(c^\dagger;e^*)-\int_0^{c^\dagger} cdF_2(c;e^*)\right\}=\gamma'(q^*+e^*). \tag{30}$$

Recall that the first best is defined by (4) and (5) where $c(e)$ is now the expected value of c given e. After applying these to the above first-order conditions, and simplifying, I obtain:

$$pF(c^\dagger;e^*)+\int_{c^\dagger}^{\hat{c}}(m+c)dF(c;e^*)=(1-\theta)\frac{W'(q^*)}{\mu'(q^*)} \tag{31}$$

$$pF_2(c^\dagger;e^*)-\int_{c^\dagger}^{\hat{c}}(m+c)dF_2(c;e^*)=0. \tag{32}$$

Together with the no dumping condition:

$$c^\dagger \le p+\theta\frac{W(q^*)}{\mu(q^*)},$$

equations (31) and (32) are the necessary conditions for the mixed system $(p, c^\dagger, m)$ to implement the first best.

The necessary conditions can be satisfied if cost-reduction efforts only affect the distribution of less severe cases, so that for c sufficiently close to the upper end of the support $(\hat{c})$, $F_e(c;e)=0$. Then any p and m will satisfy (32) when $c^\dagger$ is sufficiently large. The policy variables p and m can be chosen to satisfy (31) as well as the no dumping condition (27).

7 Concluding Remarks

I have analyzed the incentive effects of payment systems when a provider is partially altruistic. The altruism assumption seems particularly realistic in the health care industry. From a theoretical point of view, the assumption introduces another way in which the provider's preferences can be misaligned with the social objective. I find that the first best can be implemented by a prospective payment system when an altruistic provider cannot refuse supplying treatments. The effect of altruism mainly lowers the prospective price level. When the degree of altruism is sufficiently high, the provider does not make any profit, and the prospective payment resembles an average-cost pricing rule.

Cost reimbursements do not supply any cost-reduction incentives. But the margin over costs can induce quality efforts. Again, provider altruism implies a lower margin and profit. As the degree of altruism increases, the margin just pays for the disutility of the quality effort, profit becomes zero, and average cost of each treated patient is reimbursed.

When the provide may refuse treating some patients, altruism actually may alleviate some problem due to dumping because the threshold of cost at which dumping occurs is higher than the level of prospective payment. Nevertheless, for the implementation of the first best, the provider's preferences must be completely aligned with the social objective. So two set of sufficient conditions for the implementation of the first best by a pure prospective payment system in situations where dumping may or may not be used turn out to be exactly identical. I also derive a set of necessary conditions for the implementation of the first best under a combination of prospective payment and cost reimbursement.

The analysis here extends previous results by allowing the provider to be partially altruistic. From a policy view point, provider altruism is beneficial for incentive reasons. First, prospective payment levels and margins over costs will be lower when the provider is more altruistic. This means lower profit levels to the provider no matter which type of payment is used. Second, even when dumping is possible, its occurrence becomes less frequent. My model is limited to the analysis of a single provider. Competition among providers when they are altruistic is important and awaits future research.

References

Allen, Robin and Paul Gertler (1991). "Regulation and the Provision of Quality to Heterogenous Consumers: The Case of Prospective Pricing of Medical Services," *Journal of Regulatory Economics*, 3, 361-375.

Arrow, Kenneth J. (1963). "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53, 941-969.

Chalkley, Martin and Malcomson, James M. (1995). "Contracting for Health Services with Unmonitored Quality," Discussion Papers, University of Southampton.

_____ (1998). "Contracting for Health Services when Patient Demand Does Not Reflect Quality," *Journal of Health Economics*, forthcoming.

Dranove, David (1987). "Rate-setting by Diagnosis Related Groups and Hospital Specialization," *Rand Journal of Economics*, 18, 417-427.

Eisenberg, J. M. (1986). *Doctors' Decisions and the Cost of Medical Care*, Health Administration Press, Ann Arbor, MI.

Ellis, Randall P. (1998). "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins" *Journal of Health Economics*, forthcoming.

_____ and Thomas G. McGuire (1986). "Provider Behavior under Prospective Reimbursement," *Journal of Health Economics*, 5, 129-151.

Ma, Ching-to Albert (1994). "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics & Management Strategy*, 3, 93-112..

Newhouse, Joseph P. (1983). "Two Prospective Difficulties with Prospective Payment, or, Its Better to be a Resident than a Patient with a Complex Problem," *Journal of Health Economics*, 2, 269-274.

_____ (1984). "Cream Skimming, Asymmetric Information, and a Competitive Insurance Market," *Journal of Health Economics*, 3, 97-100.

Pauly, Mark (1984). "Is Cream Skimming a Problem for the Competitive Medical Market?" *Journal of Health Economics*, 3, 87-95.

Pope, Gregory C. (1989). "Hospital Nonprice Competition and Medicare Reimbursement Policy," *Journal of Health Economics*, 8, 147-172.

Russell, Louise B. (1989). *Medicare's New Hospital Payment System*, The Brookings Institution, Washington D.C.

Shleifer, Andrei (1985). "A Theory of Yardstick Competition," *Rand Journal of Economics*, 16, 319-327.