Health-Care Payment Systems: Cost and Quality Incentives—Reply

Ching-to Albert Ma
Department of Economics
Boston University
Boston, MA 02215
ma@econ.bu.edu

I reconsider the implementation of efficient cost and quality efforts when health-care providers may refuse services to consumers, and introduce a mechanism that is a combination of prospective payment and cost reimbursement. Conditions are derived for the prospective payment level and the margin above cost reimbursement for the implementation of efficient efforts.

1. Introduction

In an earlier paper (Ma, 1994), I considered the implications of prospective payment and cost reimbursement systems on provider quality and cost reduction incentives in the health market. Patient dumping, in which costly patients are denied services by providers, was one of the issues discussed in that paper. Because pure prospective payment is a fixed-price mechanism, providers will suffer a loss giving services at a cost higher than the payment level. I attempted to show in my earlier paper that a combination of prospective payment and cost reimbursement could solve the dumping problem and simultaneously provide incentive for efficient quality and cost reduction efforts. As Sharma (1998) has shown, the argument there was flawed. In this note, I present a modification of the "mixed" prospective-reimbursement mechanism, and conditions for the implementation of efficient incentives.

I begin by setting up the model. A health-care provider can expend *effort* to lower its treatment costs or enhance care quality. The variables t_1 and t_2 denote efforts a provider can direct to quality enhancement and cost reduction dimensions respectively.[1] Efforts impose a total disutility of $\gamma(t_1 + t_2)$ to the provider; the function γ is increasing and convex. As in my earlier paper, I will also use t_1 to denote the quality of care. The increasing and concave function $\mu(t_1)$ is the pro-

1. Minimum effort levels are normalized at zero.

Journal of Economics & Management Strategy, Volume 7, Number 1, Spring 1998, 139–142

vider's demand[2] when quality of service is t_1. The hospital's cost of treating a patient, c, is distributed according to some (cumulative) distribution $F(c, t_2)$ and density function $f(c, t_2)$ on the support $[0, \hat{c}]$. The function F is assumed to be continuously differentiable. Let $\bar{c}(t_2)$ denote the average cost of treatment if the hospital chooses effort t_2 (the expectation of c given t_2). I assume that the function $\bar{c}(t_2)$ is decreasing in t_2, so that higher effort leads to a cost improvement.[3]

It was shown in Ma (1994) that absent dumping, a pure prospective payment mechanism can implement the efficient mix of quality and cost efforts; let p denote the corresponding efficient prospective payment level.[4] With dumping, the provider may not internalize all costs, picking a suboptimal cost reduction effort. Earlier I proposed the following mechanism: the payment to the provider is $c^{\dagger}$ for $c \le c^{\dagger}$, and c otherwise (see Ma, 1994, pp. 106–107). The error of the claim that this mechanism avoided dumping while implementing efficient efforts lies in the neglect of the provider's optimization conditions. Sharma found that, given my mechanism, the provider may not always find it profitable to exert the efficient efforts.

2. The Implementation of Efficient Efforts

It is useful to begin by writing down the provider's profit under prospective payment when dumping is impossible:

$$\mu(t_1) \int_0^{\hat{c}} (p - c) f(c, t_2)\, dc - \gamma(t_1 + t_2).$$

Assume that the profit function is concave in (t_1, t_2). The necessary and sufficient first-order conditions with respect to t_1 and t_2 are

$$\mu'(t_1) \left(p - \int_0^{\hat{c}} c f(c, t_2)\, dc \right) = \gamma'(t_1 + t_2), \tag{1}$$

$$-\mu(t_1) \int_0^{\hat{c}} c f_{t_2}(c, t_2)\, dc = \gamma'(t_1 + t_2). \tag{2}$$

Again, let p be set at the level for implementing the efficient efforts t_1^* and t_2^*.

2. Consumers are assumed to have zero deductible or coinsurance.

3. This may be due to an improvement of cost in the sense of first-order stochastic dominance, but such an assumption is not made here or in my earlier paper.

4. The value of p is given by $W'(t_1^*)/\mu'(t_1^*)$, where W is consumer benefit and t_i^* are the efficient efforts; see equation (10) in Ma (1994).

I now define a mixed prospective-cost reimbursement system. Such a payment mechanism is defined by a pair (c^*, m): if the provider incurs a cost c below c^*, it will be paid c^*; above c^*, $c + m$, where $m \geq 0$. Thus, for cost realizations below c^*, the provider receives a fixed price c^*; above c^*, a constant margin m over cost. This mixed system consists of *two* instruments; in Ma (1994) as well as Sharma (1998), m is set at zero. Clearly, the pure prospective and cost reimbursement systems are special cases (respectively, for $c^* = \hat{c}$ and $c^* = 0$). Because net payment is always above cost, dumping is not profitable, and the firm's profit becomes

$$\mu(t_1)\left(\int_0^{c^*} (c^* - c)f(c, t_2)\, dc + \int_{c^*}^{\hat{c}} mf(c, t_2)\, dc\right) - \gamma(t_1 + t_2).$$

The first-order conditions with respect to t_1 and t_2 are

$$\mu'(t_1)\left(\int_0^{c^*} c^*f(c, t_2)\, dc + \int_{c^*}^{\hat{c}} (m + c)f(c, t_2)\, dc - \int_0^{\hat{c}} cf(c, t_2)\, dc\right) = \gamma'(t_1 + t_2), \tag{3}$$

$$\mu(t_1)\left(\int_0^{c^*} c^*f_{t_2}(c, t_2)\, dc + \int_{c^*}^{\hat{c}} (m + c)f_{t_2}(c, t_2)\, dc - \int_0^{\hat{c}} cf_{t_2}(c, t_2)\, dc\right) = \gamma'(t_1 + t_2). \tag{4}$$

I assume that the first-order conditions are necessary and sufficient.

The first-best efforts [as given by the first order conditions (1) and (2)] are implementable if and only if there are c^* and m such that equations (3) and (4) yield the first-best efforts. After simplification, this requirement becomes

$$p = \int_0^{c^*} c^*f(c, t_2)\, dc + \int_{c^*}^{\hat{c}} (m + c)f(c, t_2)\, dc, \tag{5}$$

$$\int_0^{c^*} c^*f_{t_2}(c, t_2)\, dc + \int_{c^*}^{\hat{c}} (m + c)f_{t_2}(c, t_2)\, dc = 0. \tag{6}$$

Proposition 1: *The first-best efforts, t_1^* and t_2^*, are implementable if and only if there exist c^*, $0 \leq c^* \leq \hat{c}$, and $m \geq 0$ such that equations* (5) *and* (6) *are satisfied for $t_i = t_i^*$, $i = 1, 2$.*

Applying integration by parts, and by the assumption that F is continuously differentiable [so that $F_{t_2}(0, t_2) = F_{t_2}(\hat{c}, t_2) = 0$], I rewrite equation (6) as

$$-mF_{t_2}(c^*, t_2) - \int_{c^*}^{\hat{c}} F_{t_2}(c, t_2)\, dc = 0, \tag{7}$$

which therefore is a necessary condition for implementation (at $t_2 = t_2^*$). Notice that equation (7) is *not* inconsistent with the usual definition of first-order stochastic dominance, which in our context is $F_{t_2}(c, t_2) \geq 0$ with strict inequality for at least some c. Equation (7) can be satisfied when $F_{t_2}(c, t_2) = 0$ for all c sufficiently close to $\hat{c}$, since then c^* can be chosen to be sufficiently close to $\hat{c}$. [Subsequently, an appropriate choice of m will satisfy equation (5).]

Sharma (1998) has pointed out correctly that if $F_{t_2} > 0$ for all $0 < c < \hat{c}$ and t_2—an increase in t_2 improves costs in the sense of *strict* first-order stochastic dominance—then equation (7) will be impossible to satisfy, and efficient efforts cannot be implemented by a combination of prospective payment and cost reimbursement. Of course, strict dominance is a more stringent assumption than the usual notion of dominance. In any case, neither in my earlier paper nor here have I used an assumption of stochastic dominance. Furthermore, because stochastic dominance is a partial ordering, but the comparison of expected values of two distributions can be made whenever they exist, my assumptions that $\bar{c}(t_2)$ is decreasing in t_2 is more general. In addition, because Sharma did not allow a margin with cost reimbursement, his condition (6) in fact is *not* necessary for implementation; for the same reason, the situation illustrated by his Figure 2 need not be satisfied for the implementation of efficient efforts.

A combination of prospective payment and cost reimbursement must perform better than a fixed-price or cost-plus system alone—trivially, it includes each of the constituent systems as a special case. The best way such a system should be used is less clear. Here, I present conditions so that the mixed mechanism can maintain cost and quality incentives and avoid dumping. Sharma (1998) has shown that it may provide cost incentive as well as satisfy a distributional trade-off between profits and incentives. The full analysis of the incentive properties of a mixed system awaits future research.

References

Ma, C.A., 1994, "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics & Management Strategy*, 3(1), 93–112.

Sharma, R.L., 1998, "Health Care Payment Systems: Cost and Quality Incentives—Comment," *Journal of Economics and Management Strategy*, 7(1), 127–137.