Abstract

A health care provider chooses medical service quality and cost-reduction effort. Both choices are non-contractible. An insurer observes both quality and cost effort, and may credibly disclose them to consumers. In prospective payment, the insurer fully discloses care quality, and sets a prospective payment price. In cost reimbursement, the insurer discloses a value index, a weighted average of quality and cost effort, and pays a margin above cost. The first-best quality and cost effort can be implemented by prospective payment and by cost reimbursement.

Keywords: information disclosure, prospective payment, cost reimbursement, fee for service, quality, cost reduction

Acknowledgement: We have had many inspiring conversations with Mike Luca. For their comments, we thank Chiara Canta, Randy Ellis, Iris Kesternich, Ting Liu, Tom McGuire, Sara Machado, Inés Macho Stadler, Pau Olivella, Alex Poterack, Alyson Price, Jim Rebitzer, Sam Richardson, Tomás Rodríguez Barraquer, and seminar participants at Boston University, European University Institute, Harvard University, Lehigh University, University of Salerno, and the 13th European Health Economic Workshop. The second author received financial and research support from the Max Weber Programme at European University Institute.
1 Introduction

The (provocative) title refers to prospective payment and cost reimbursement, the most common mechanisms for paying health care providers. In prospective payment, a health care provider receives a fixed price for delivering a course of medical services, irrespective of the resources that have been used. In cost reimbursement, a provider receives a measure of costs corresponding to the resources for the delivery of medical services. These two payment methods have been studied extensively and intensively in the past thirty years. The conventional wisdom is that prospective payment and cost reimbursement give rise to different quality and cost-efficiency incentives. In this paper, we describe a model in which prospective payment and cost reimbursement can give rise to identical quality and cost incentives. This model actually differs from the conventional one only in how consumers learn about quality.

The canonical model for studying quality and cost efficiency is this. A health care provider chooses quality and cost-reduction efforts. These efforts are noncontractible. The provider incurs private disutilities by expending these efforts. Costs of providing services consist of a marginal cost as well as the effort disutilities. A higher quality requires a higher marginal cost and attracts more consumers, but a higher cost effort reduces the marginal cost. An insurer wants to implement socially efficient quality and cost efforts.

Under prospective payment, the provider internalizes the production cost, so its cost-reduction incentive is aligned with the insurer’s. What about the provider’s quality incentive? Seeking to maximize profit, the provider considers raising quality to attract more consumers. The marginal benefit of this depends on the prospective payment level, and this has to be traded off against the quality effort marginal disutility. By choosing the appropriate prospective payment level, the insurer aligns the provider’s profit motive to one consistent with the implementation of the socially efficient quality. Prospective payment can kill two birds with one stone.

Cost reimbursement works in a perverse way. Because all marginal costs will be reimbursed, the provider

---

For our purpose, cost reimbursement is the same as fee-for-service: a provider chooses medical services to supply, and receives a fee for each chosen service. This fee reflects the cost of the service and allows a profit margin. There are variations in prospective payment; it may be supplemented by outlier compensations, local-market adjustments, etc. These variations are unimportant for this paper.
lacks any incentive to use cost effort. Cost reimbursement results in cost inefficiency. The quality incentive can still be implemented by paying the provider a margin above cost for services rendered. Again, the provider raises quality to attract more consumers because of the profitable margin.

In the two payment systems, the common principle motivating quality is demand response: higher quality raises demand. In each system, a profit margin incentivizes the provider to expend quality effort. The difference is that the provider internalizes costs under prospective payment, but does not do so under cost reimbursement.

The notion of a demand response requires consumers to know about quality. This is a common assumption in the literature (see more discussion below). Naturally, if consumers could never be conveyed service quality information, the health care market would collapse completely. Given that the health market is very active in all economies, some quality information must be available to consumers. Most health economists, however, would agree that health care quality information can be difficult to obtain and interpret. Indeed, insurers, governments, and sponsors increasingly have helped consumers find out about quality. For a summary of empirical works on public reporting initiatives, see Dranove and Jin (2011).

In this paper, we make an alternative assumption about information structures in the canonical model. Here, we assume that consumers cannot observe quality directly, but the insurer can. Furthermore, although cost effort is noncontractible, the insurer can also observe it. The insurer can credibly disclose information about quality and cost efforts to consumers. The canonical model obtains if the insurer simply discloses any quality effort chosen by the provider. (Consumers are insured and uninterested in cost effort anyway.) In fact, if a prospective payment system is used, the insurer simply fully discloses quality information, and implements the first-best quality and cost efforts.

The surprise is that the insurer can use cost reimbursement to implement the first best by disclosing partial information about quality and cost. Our innovation is that by disclosing a value index, a weighted average of quality and cost efforts, the insurer incentivizes the provider to undertake cost effort, even when all costs are reimbursed. The key is that demand depends on consumers’ perception about quality from the value index. The insurer insists on mixing quality and cost-effort information in the quality index. The
incentive for cost effort comes from the provider achieving a value index by profit-maximizing quality and cost effort. Indeed, cost reimbursement and strategic information disclosure implement the same allocation as prospective payment. Before we explain this result, we should point out its relevance.

Prospective payment has various unintended consequences. First, because the price is fixed, the provider takes a loss when treating high-cost consumers. Second, for the same reason, the provider earns more profit by attracting low-cost consumers. Dumping and cream-skimming under prospective payment have been studied extensively in the literature. Third, prospective payment encourages fraudulent upcoding. For hospitals, in the actual implementation of prospective payment, the Diagnostic Related Group system is used: after the treatment episode, the provider reports the consumer’s primary diagnosis for payment. The so-called DRG creep refers to a provider gaming the insurer by misreporting a consumer’s diagnosis to get a higher price. For physicians, prospective payment encourages seeking a higher price by lying about the actual treatment.

The current theoretical and policy debates have been heavily against cost-based payments. Cost reimbursement has none of the problems of dumping, cream skimming, and upcoding, simply because under cost reimbursement, consumer cost heterogeneity is of no concern to the provider. Cost reimbursement avoids a host of selection issues. The current sentiment is that cost reimbursement is a bad policy because of cost inefficiency. In this paper, we show how cost efficiency can be made consistent with cost reimbursement.

The incentive mechanism for the provider to exert cost effort under cost reimbursement works as follows. We assume that quality is not observed by consumers, but an insurer can observe both quality and cost efforts. Our innovation is to let the insurer construct a value index—a weighted average of the quality and cost efforts—and disclose this to consumers.

Still, why would the provider exert cost effort under cost reimbursement? Consumers only observe the value index, not quality, so they will infer about quality based on the value index. A given level of value index therefore corresponds to some inferred quality level, generates a demand, and, hence, profits. However, the insurer mixes quality and cost effort to construct the value index. The provider could invest in quality alone to achieve any value index, but it would get a higher profit by investing in cost effort also.
For example, suppose that the value index puts equal weights on quality and cost effort. To achieve a value index of 100 by quality alone, quality would have to be 200 (which yields $100 = 200 \times 0.5 + 0 \times 0.5$). However, the provider could achieve that index by choosing both quality and cost effort at 100. Quality and cost effort generate disutilities, so among the many combinations of quality and cost effort that can generate a value index, the provider will choose the profit-maximizing one. Generally, the profit-maximizing cost effort is strictly positive. Furthermore, the insurer can choose the index weight and profit margin to make the provider internalize the net social benefit of cost effort and quality.

It has not escaped our notice that our theory relies on the provider being unable to disclose credibly quality information. If a provider was able to do so, it could defeat the value-index manipulation. In practice, there does not seem to be any “danger” that any provider could fully disclose quality information. Otherwise, public agencies (such as the Center for Medicare and Medicaid Services) and nonprofit organizations (such as Consumer Reports and the National Committee for Quality Assurance) would not have expended huge resources on quality reports to the general public. Furthermore, it is far from clear that a provider would report honestly quality information even when it was feasible to do so.

The literature on provider payment design is large. For extensive surveys of theoretical and empirical findings, see Newhouse (1996), McGuire (2000), and Leger (2008). Ma (1994) laid out the basic model of health care payment systems and their effects on quality and cost incentives. The general consensus is that cost reimbursement fails to achieve cost efficiency, and that prospective payment leads to perverse selection incentives such as dumping and creaming. Generally neither cost reimbursement nor prospective payment achieves socially efficient outcomes.

In recent years, many insurers have introduced reforms to complement cost reimbursement and prospective payment (McClellan, 2011). These payment schemes tend to be a mix of prospective payment and cost reimbursement, as well as new elements such as pay for performance, and ex post risk-adjusted payments. This paper keeps prospective payment and cost reimbursement at their simplest forms, and focuses on how an insurer can use quality and cost-effort information disclosure to incentivize providers. This in turn allows us to offer a mechanism of information disclosure and cost reimbursement to resolve the trade-off between
cost efficiency and selections.

We assume a demand response: consumers’ demand for services reacts positively to quality. This is an assumption that is almost universally adopted in the literature: see for example, Rogerson (1994), Ma and McGuire (1997), Frank et al. (2000), Glazer and McGuire (2000), Brekke et al. (2006). One exception in the payment design literature is Chalkley and Malcomson (1998); they posit that even when quality increases, more demand cannot be satisfied due to limited capacities and rationing, common in many European systems. Chalkley and Malcomson then assume that the provider is altruistic. Altruism motivates quality efforts. We use a conventional assumption that the provider seeks to maximize profits.

A number of recent papers empirically evaluate demand response to public reports. In commercial health-plan markets, both Beaulieu (2002) and Scanlon et al. (2002) show that consumers do avoid health plans with low ratings. Since 1999, the Center for Medicare and Medicaid Services has launched quality-report initiatives for health plans, hospitals, physicians, and nursing homes (see www.cms.gov/QualityInitiativesGenInfo/). Dafny and Dranove (2008) find that the reports for Medicare health plans substantially affected enrollments.

Our paper is closely related to a small but growing literature on optimal public-report design. Glazer and McGuire (2006) propose that a regulator can solve an adverse-selection problem in a competitive market by reporting only average-quality information. Their concern is quality for \textit{ex ante} heterogenous consumers, and their mechanism achieves cross subsidies among consumers and first-best qualities by average-quality reports. Ma and Mak (2012) characterize the optimal average-quality reports that mitigate monopoly price discrimination and quality distortion. The current paper contributes to the literature by simultaneously studying optimal payment and reporting policies. In particular, we show that an optimal reporting policy can induce socially efficient cost effort under cost reimbursement.

Information asymmetry has long been viewed as a source of inefficiency in the physician-patient interaction literature. For example, in both Dranove (1988) and Rochaix (1989), a physician utilizes his private information to induce patient demand for excessive treatments. Instead, the insurer in our model holds back some information from consumers to induce cost-reduction effort. Our main result shows that information asymmetry improves efficiency in physician-patient interaction under cost reimbursement.
Information disclosure has been extensively studied in the industrial organization literature. In Matthews and Postlewaite (1985) and Schlee (1996), product quality is unknown to the seller, consumers, or both. They show that quality information can harm consumers because of the seller’s price response. We instead focus on how a trusted intermediary can utilize demand response to discipline a seller. In both Lizzeri (1999) and Albano and Lizzeri (2001), a profit-maximizing intermediary privately observes product quality. They show that the intermediary may underprovide quality information at the expense of market efficiency. But the insurer in our model withholds information to achieve efficient quality and cost effort.

The plan of the paper is as follows. Section 2 presents the model. Section 3 sets up the information structure, and derives our main result. Section 4 considers four robustness issues. We first show the implementation of the first best i) when consumers may misinterpret the value index, ii) when cost information, rather than effort, is observed, and iii) when the provider chooses many qualities. Then we show that cost reimbursement outperforms prospective payment when a provider can practice dumping and cream-skimming. Finally, Section 5 draws some conclusions.

2 The Model

2.1 Consumers and a provider

A set of consumers are covered by an insurer. We let the insurance coverage be complete, so consumers have no copayments. Health services are to be supplied by a provider. If consumers believe that health care quality is $q$, the quantity demanded is $D(q)$. The function $D$ is strictly increasing and concave. The social benefit from quality $q$ is denoted by $B(q)$ where $B$ is a strictly increasing and concave function. In many applications $B$ is consumer benefit from services, but we allow a more general interpretation so that externalities, equity, and any other such issues can be included.

A provider supplies health services to the insured consumers. It chooses the quality of care $q$ and a cost-reduction effort $e$, both nonnegative. The unit cost for service is $C(q, e)$ when the provider chooses quality $q$ and effort $e$. The function $C$ is strictly increasing in $q$ and strictly decreasing in $e$, and convex. A
higher quality of care requires a higher cost, but cost can be reduced by the provider’s effort. In addition, the provider incurs two fixed costs or disutilities for quality and effort, namely \( G(q) \) and \( H(e) \). The two functions \( G \) and \( H \) are strictly increasing and convex. With quality \( q \) and effort \( e \), the demand will be \( D(q) \), so the provider incurs a total cost \( D(q)C(q, e) + G(q) + H(e) \).

### 2.2 Payment mechanisms and information

Quality and cost-reduction effort are noncontractible. The quantity of services is observed \textit{ex post} and payment can be based on it. The unit cost of services \( C(q, e) \) is also \textit{ex post} observed, and again payment can be based on it. The fixed cost of quality and the disutility of cost effort are unobservable. These are standard assumptions reflecting the complexity of quality and effort, as well as common payment policies.

In the literature, two forms of payments have been extensively studied: prospective payment and cost reimbursement. Prospective payment is a fixed price \( p \) per unit of delivered service. Under cost reimbursement, the provider will be paid the variable cost \( C(q, e) \) plus a margin \( m \) per unit of delivered services. Prospective payment \( p \) and the margin \( m \) are nonnegative. We will study these two forms of payment. We also include a lump-sum payment, a transfer, for the provider. This transfer can be positive or negative.

Our departure from the classical payment-design problem is on the information about quality and cost effort. In the literature, consumers are assumed to observe the quality \( q \) and their demand for services is straightforwardly given by the demand function \( D(q) \). We consider an alternative scenario. Here, consumers are unable to observe quality and cost effort directly. This actually is consistent with the maintained assumption that quality and effort are complex and noncontractible, and arguably more realistic.

The insurer acts as a trusted information intermediary. The insurer observes the provider’s choice of quality and effort. The insurer then decides whether to disclose this information. The insurer may disclose information fully. The insurer may also choose to disclose an index. If quality \( q \) and effort \( e \) have been chosen, the insurer may construct a weighted average \( I(\theta) = \theta q + (1 - \theta)e \), where \( 0 \leq \theta \leq 1 \). This index is then reported to consumers. We will call \( I(\theta) \) a \textit{value index}.

---

2 The model can be easily extended to incorporate cost heterogeneity. Dumping and cream skimming for the current model have been addressed by Ma (1994). More discussions are in Subsection 4.4.
Consumers are interested in quality. Their demand for health services does not depend on cost effort. If we set the weight of the value index $\theta$ to 1, then full quality information will be revealed to consumers. If $\theta$ is always set to 1, consumers observe the provider’s quality choice and respond by demanding health care; this would be the standard model. The point of our paper, however, is that the weight should be set different from 1 under cost reimbursement.

### 2.3 The first best

In the first best, quality and cost effort are contractible. The social welfare from quality $q$ and effort $e$ is

$$B(q) - D(q)C(q, e) - G(q) - H(e),$$

which is simply the social benefit less the total cost. Let $q^*$ and $e^*$ be the quality and effort that maximize social welfare in (1). They are characterized by the first-order conditions:

$$B'(q^*) - D'(q^*)C(q^*, e^*) - D(q^*)C_q(q^*, e^*) - G'(q^*) = 0$$

$$-D(q^*)C_e(q^*, e^*) - H'(e^*) = 0,$$

where we use the usual notation to denote derivatives and partial derivatives. The first-order conditions have the usual marginal interpretations. Raising quality increases social benefit, but it also raises demand (hence total cost), unit cost, and fixed cost. Raising cost effort reduces unit cost but raises fixed cost. The first-order conditions in (2) and (3) balance these marginal effects.

### 3 Prospective payment, cost reimbursement, and value index

#### 3.1 Prospective payment and first best

We let the insurer either operate in a competitive market, or be a public agency. The insurer’s objective is to maximize a weighted sum of social net benefit and the provider’s profit, with a lower weight on profit.

In a prospective payment system, the provider is paid a fixed price $p$ per unit of service, together with a transfer $T$. Suppose that the insurer fully discloses quality $q$ ($\theta = 1$). When the provider chooses quality $q$ and effort $e$, its payoff is

$$T + pD(q) - D(q)C(q, e) - G(q) - H(e).$$

(4)
The quality and cost effort generate a social net benefit

\[ B(q) - pD(q) - T, \]  

which is the social benefit \( B(q) \) less payments to the provider.

The insurer’s objective is to choose the prospective price \( p \) and the transfer \( T \) to maximize

\[ w[B(q) - pD(q) - T] + (1 - w)[T + pD(q) - D(q)C(q, e) - G(q) - H(e)], \]  

where \( .5 < w \leq 1 \). The provider must make a nonnegative profit, so the expression in (4) must be nonnegative. Given that the welfare weight is larger on social net benefit, the optimal transfer \( T^* \) will make sure that profit in (4) is exactly zero. The insurer’s objective is then simplified into a choice of price \( p \) to maximize (5). A choice of \( p \) implements the provider’s best response of choosing \( q \) and \( e \) to maximize its profit (4).

The following proposition is adapted from Ma (1994), and stated with its proof omitted:

**Proposition 1**: By choosing \( p^* = \frac{B'(q^*)}{D'(q^*)} \) and a suitable transfer \( T^* \), the insurer implements the first-best quality \( q^* \) and cost effort \( e^* \).

The intuition is well documented in the literature. Under prospective payment, the provider fully internalizes the social cost of quality and cost-reduction effort. Its incentive on cost efficiency aligns with the insurer’s. By setting the prospective price at the \( p^* \) in Proposition 1, the insurer makes the provider internalize the social benefit of quality as well. Any profit from the prospective payment is taxed away by the transfer, so the first best is implemented.

### 3.2 Cost reimbursement, value index, and first best

Under cost reimbursement, the insurer commits to reimburse the provider's variable cost, and pays a margin \( m \) for each unit of delivered services. Furthermore, the insurer will report on the provider’s choices of quality and cost effort in the form of a value index. We study the perfect-Bayesian equilibria of the following extensive-form game:

**Stage 1**: The insurer sets the transfer \( T \), the margin \( m \), and the weight \( \theta \) in the value index. The insurer also commits to reimbursing the provider’s operating cost.
**Stage 2:** The provider chooses quality $q$ and effort $e$.

**Stage 3:** The insurer observes the provider’s choices of quality $q$ and effort $e$, and reports the value index $I(\theta) = \theta q + (1 - \theta) e$ to consumers.

**Stage 4:** Consumers learn $I(\theta)$ (but never the provider’s choices of $q$ and $e$), and decide on the quantity of services to obtain.

Consumers do not observe the provider’s quality choice, and must infer it from the value index, so we consider perfect-Bayesian equilibria. What are consumers’ equilibrium beliefs? Suppose that in an equilibrium, the provider chooses quality $\hat{q}$ and $\hat{e}$, and therefore the value index is $I(\theta) = \theta \hat{q} + (1 - \theta) \hat{e}$. Then in equilibrium, consumers must believe quality to be $\hat{q}$, and their demand will be $D(\hat{q})$. Given this belief, the provider’s profit from choosing any quality $q$ and effort $e$ satisfying $\theta q + (1 - \theta) e = \theta \hat{q} + (1 - \theta) \hat{e}$ is

$$T + mD(\hat{q}) - G(q) - H(e). \quad (7)$$

Any change of $(q, e)$ from $(\hat{q}, \hat{e})$ cannot be detected by consumers as long as they generate the same value index $\theta q + (1 - \theta) e = \theta \hat{q} + (1 - \theta) \hat{e}$, so equilibrium quality $\hat{q}$ and effort $\hat{e}$ must maximize profit.

**Lemma 1 :** Equilibrium quality and cost effort $(\hat{q}, \hat{e})$ must solve

$$\max_{q,e} T + mD(\hat{q}) - G(q) - H(e)$$

subject to $\theta q + (1 - \theta) e = \theta \hat{q} + (1 - \theta) \hat{e}$.

Hence $(\hat{q}, \hat{e})$ satisfy

$$\frac{G'(\hat{q})}{H'(\hat{e})} = \frac{\theta}{1 - \theta}. \quad (8)$$

Lemma 1 says that equilibrium quality and cost effort must minimize their combined fixed cost $G$ and $H$ to achieve any level of the value index.\(^3\) The condition in (8) gives the optimality condition for the

---

\(^3\)The more theoretically inclined reader must notice that Lemma 1 also specifies consumer beliefs off the equilibrium path. The Lemma says that for any quality index, not just the one chosen by the provider in equilibrium, consumers believe that quality and cost effort have been chosen to minimize the disutility. This can be justified by a weak belief restriction. The provider’s strategy of choosing quality and cost that do not minimize disutility for some value index is dominated by one that does. Therefore, Lemma 1 essentially says that consumers never believe that the provider chooses a weakly dominated strategy.
minimization of $G(q) + H(e)$ subject to $\theta q + (1 - \theta)e$ being set at some fixed level. The ratio of the marginal disutilities $G'(q)/H'(e)$ must be equal to the ratio of the quality and cost weights $\theta/(1 - \theta)$.

Even when unit variable costs, $C(q, e)$, are completely reimbursed, the provider still has an incentive to put in cost-reduction efforts. The key is that consumers respond to quality, but they only observe the value index, so they infer quality from the value index. If the insurer makes both quality and cost effort contribute to the quality index by setting $\theta$ between 0 and 1, incentives for cost effort are feasible.

Figure 1 illustrates Lemma 1. The downward sloping straight line is an iso-index line. It plots the combination of quality and cost effort that give rise to a certain value of the index, say $I_1(\theta)$. The iso-disutility line, which is concave to the origin, describes those $(q, e)$ pairs that yield a constant disutility $G(q) + H(e)$. The tangency point $(\tilde{q}_1(\theta), \tilde{e}_1(\theta))$ minimizes disutility subject to the value index constraint.

The provider can choose various combinations of quality and effort to achieve different levels of the value index. As the level of the value index changes, different tangency points result: the “expansion path” is the upward-sloping dotted line. A change in the value of the weight $\theta$ will tilt the expansion path. For example, if $\theta$ increases so that quality has a higher weight in the index, the provider will choose more quality and less effort. This corresponds to the iso-index line being pivoted in a clockwise direction.
Given a margin $m$ and a value-index weight, the provider’s equilibrium quality $q$ and $e$ are those that maximize $T + mD(q) - G(q) - H(e)$ subject to (8). We assume that this constrained optimization problem is well-behaved, so the first-order conditions are necessary and sufficient. The following proposition says that cost reimbursement with value-index reporting achieves the first best.

**Proposition 2 :** By choosing $\theta^* = \frac{G'(q^*)}{G''(q^*) + H'(e^*)}$, $m^* = \frac{G'(q^*)}{D'(q^*)} + \frac{H'(e^*)}{D'(q^*)} \times \frac{d\ln G'(q^*)}{d\ln H'(e^*)}$, and a suitable transfer $T$, the insurer implements the first-best quality $q^*$ and cost effort $e^*$.

By Lemma 1, the insurer can choose a weight so that the first-best quality and cost effort minimize disutility $G(q) + H(q)$. This is the weight $\theta^*$ obtained by solving for $\theta$ in (8) at $\hat{q} = q^*$ and $\hat{e} = e^*$. In other words, the weight $\theta^*$ ensures that the first best is on the expansion path, so it is a candidate for an equilibrium; see Figure 2.

---

12

---

*Use the constraint (8) to define implicitly $e$ as a function of $q$. Then substitute $e$ in the profit function, which now has a single variable $q$. This is equation (18) in the proof of Proposition 2. The provider’s first-order condition is sufficient if (18) is quasi-concave in $q$. It is straightforward to find conditions in terms of derivatives of the various functions to guarantee that (18) is concave.*
Next, the margin $m$ is to be chosen so that the first best indeed is the provider’s equilibrium choice. Respecting the constraint (8) in Lemma 1, the provider chooses between $q$ and $e$ according to the usual trade-off. Consider the iso-profit line obtained by setting $T + mD(q) - G(q) - H(e)$ to a constant. This implicitly defines a function $e$ in terms of $q$. The derivative of the iso-profit line is $[mD'(q) - G'(q)]/H'(e)$, which is positive for low $q$’s but turns negative at high $q$’s, so this function has an inverted U-shape, like the one in Figure 2. Points below the iso-profit line yield higher profits to the provider. And a higher value of $m$ shifts this function upward in the $q$-$e$ space. The equilibrium is the tangency point $(q^*, e^*)$.

We now interpret the optimal margin $m^*$. First, the constraint (8) implicitly defines the expansion path in Figure 2 as a function $e$ in terms of $q$. The slope of this function is given by $\frac{\text{d} \ln G'(q)}{\text{d} \ln H'(e)}$ (see the Appendix), which describes the proportional change of $G'$ with respect to the proportional change in $H'$. Next, equating $\frac{\text{d} \ln G'(q)}{\text{d} \ln H'(e)}$ to the slope of the iso-profit line $[mD'(q) - G'(q)]/H'(e)$ at $(q^*, e^*)$ and rearranging terms, we obtain the $m^*$ in the Proposition. In equilibrium, the $m^*$ is set such that the provider’s marginal benefit of quality investment $m^*D'(q^*)$ equates the sum of marginal disutilities $G'(q^*) + H'(e^*) \times [\text{d} \ln G'(q^*)/\text{d} \ln H'(e^*)]$.

It is important that consumers rely on the value index to infer about quality. If a provider could credibly reveal its quality, it could avoid the constraint on the equilibrium mix of quality and cost effort due to the value index (Lemma 1). Cost-effort information *per se* is not valuable to consumers. If the provider does not need to exert cost effort to convey quality information to consumers, the perverse cost effort property of cost reimbursement remains. The policy implication is perhaps quite obvious: public agencies should have a keen interest in information disclosure. A more radical policy would require public certification or regulation of any information disclosure.

A combination of value-index weight and the margin over cost reimbursement implement the first best. The literature has discussed extensively the poor cost-effort incentives under cost reimbursement, as well as the incentives for provider dumping unprofitable consumers and cream-skimming profitable ones under prospective payment. Proposition 2 offers a different perspective. Cost reimbursement eliminates incentives to dump or to cream-skim patients, but its perverse cost incentives can be avoided. We next turn to a
number of robustness issues.

4 Robustness of value index, consumer inferences, and selection

4.1 Consumer rationality

It may appear that Lemma 1 relies on consumers being fully rational. In fact, Lemma 1 stems from the firm maximizing profits. Consider an arbitrary inference rule (such as consumers naively believing that quality is always 50% of the value index). If the value index takes a value of $I$, assume that consumers believe that the quality is $\Psi(I)$, where $\Psi$ is an increasing and differentiable function.

Under cost reimbursement, given a margin $m$ and an index $I$, the provider’s profit is

$$D(\Psi(I))m - G(q) - H(e).$$

Equilibrium quality $\hat{q}$ and effort $\hat{e}$ must solve

$$\max_{q,e} D(\Psi(I))m - G(q) - H(e)$$

subject to $\theta q + (1-\theta)e = \theta \hat{q} + (1-\theta)\hat{e} = I$. Because the index $I$ is fixed in the above constrained maximization program, the first-order conditions with respect to $q$ and $e$ are:

$$-G'(q) + \lambda \theta = 0$$

$$-H'(e') + \lambda (1-\theta) = 0$$

where $\lambda$ is the multiplier. From these first-order conditions, we obtain

$$\frac{G'(q)}{H'(e')} = \frac{\theta}{1-\theta},$$

which is (8) in Lemma 1.

As long as the provider minimizes the disutilities due to quality and cost effort, the value-index weight determines how the provider must trade-off quality against cost effort, given any consumer inference. The value index therefore incentivizes the provider to reduce cost.

To ensure that the first best is an equilibrium, the insurer must set $\theta$ at $\theta^* = \frac{G'(q^*)}{G'(q^*) + H'(e^*)}$, as in Proposition 2. The implementation of the first best, however, must use a margin different from the one in
Proposition 2. Under the inference rule $\Psi$, profit is

$$D(\Psi(\theta^* q + (1 - \theta^*) c(q))) m - G(q) - H(c(q)), \quad (9)$$

where the function $c(q)$ is implicitly defined by (8) (see also the proof of Proposition 2). The monotonicity of $\Psi$ implies that there exists an $m$ such that the first-order derivative of (9) vanishes at $q = q^*$. This value of $m$ implements the first best, but generally this will be different from the one in Proposition 2.

4.2 Cost and value index

In this subsection, we change the value index into a weighted average of quality and unit cost. Using cost information may be more practical because cost may be easier to observe than effort. Now let the value index be defined by $J(\theta) = \theta q + (1 - \theta)(K - C(q, e))$, for some cost ceiling $K > 0$, and sufficiently big. Here, $K - C(q, e)$ measures the cost reduction from the preset ceiling.

Under cost reimbursement equilibrium quality and effort are those that minimize the fixed cost or disutility given any level of the value index. Hence equilibrium $\hat{q}$ and $\hat{e}$ solve

$$\min_{q, e} G(q) + H(e)$$

subject to $\theta q + (1 - \theta)(K - C(q, e)) = \theta \hat{q} + (1 - \theta)(K - C(\hat{q}, \hat{e}))$. The first-order conditions are:

$$G'(q) - \gamma [\theta - (1 - \theta)C_q(q, e)] = 0$$
$$H'(e) + \gamma(1 - \theta)C_e(q, e) = 0,$$

which simplify to

$$\frac{G'(q)}{H'(e)} = - \frac{\theta - (1 - \theta)C_q(q, e)}{(1-\theta)C_e(q, e)}.$$

(10)

The equilibrium quality and cost effort choices can be illustrated in Figure 3. Because $C$ is assumed convex, the upper contour sets of the iso-index line $\theta q + (1 - \theta)(K - C(q, e)) = J(\theta)$ are convex. In Figure 3, the iso-index lines, at index values $J_1(\theta)$ and $J_2(\theta)$, are the circular lines. The iso-disutility line is the one that is concave to the origin. An equilibrium is the tangency point between the iso-index and iso-disutility lines. Changing the index weight $\theta$ corresponds to changing the entire map of the iso-index lines. Nevertheless, for any $\theta$, (10) defines a monotone, increasing function of $e$ in $q$, say $\tilde{e}(q)$.
To implement the first best, first set $\theta$ to $\theta^{**}$ where

$$\frac{G'(q^*)}{H'(e^*)} = -\frac{\theta^{**} - (1 - \theta^{**})C_q(q^*, e^*)}{(1 - \theta^{**})C_e(q^*, e^*)}. \tag{11}$$

This guarantees that the first best is a potential equilibrium. Because (10) is linear in $\theta$, at $(q, e) = (q^*, e^*)$, there is a unique $\theta^{**}$ that satisfies (11). Next, given $\theta = \theta^{**}$, the provider’s profit is

$$mD(q) - G(q) - H(e(q)),$$

whose first-order derivative is $mD'(q) - G'(q) - H'(e(q))\tilde{e}'(q)$. We choose $m$ such that it is optimal for the provider to choose the first best:

$$mD'(q^*) - G'(q^*) - H'(e(q^*))\tilde{e}'(q^*) = 0.$$

Again, the margin that implements the first best will be different from the one in Proposition 2, but it achieves the same outcome.

### 4.3 Multiple qualities

Suppose now there are two service qualities, $q_1$ and $q_2$. We extend the definitions of demand, social benefit, cost, and disutilities in the obvious way: $D(q_1, q_2)$, $B(q_1, q_2)$, $C(q_1, q_2, e)$, $G_1(q_1)$, $G_2(q_2)$, and $H(e)$. We also
maintain the corresponding concavity and convexity assumptions.

The social welfare is now

$$B(q_1, q_2) - D(q_1, q_2)C(q_1, q_2, e) - G_1(q_1) - G_2(q_2) - H(e). \quad (12)$$

Let $q_1^*, q_2^*$, and $e^*$ be the first-best qualities and effort.\(^5\) Under prospective payment with transfer $T$ and price $p$, and complete quality-information disclosure, the provider’s profit is

$$T + pD(q_1, q_2) - D(q_1, q_2)C(q_1, q_2, e) - G_1(q_1) - G_2(q_2) - H(e).$$

If the insurer discloses information of both $q_1$ and $q_2$, a prospective price can be chosen to implement the first best if and only if

$$\frac{B_{q_1}(q_1^*, q_2^*)}{D_{q_1}(q_1^*, q_2^*)} = \frac{B_{q_2}(q_1^*, q_2^*)}{D_{q_2}(q_1^*, q_2^*)}, \quad (13)$$

(which is also the prospective price). This result is obtained by comparing the first-order conditions for the first best and for the provider’s profit maximization (as in Proposition 1).

With a single quality, a single prospective price is sufficient for the first best, as in Proposition 1. With multiple qualities, a single prospective price is insufficient generally. The provider internalizes cost under prospective payment, but profit maximization is achieved only if the (net) marginal contributions of qualities to profits are equalized. This profit-maximizing marginal contribution is generally different from each quality’s marginal contribution to social benefit. Condition (13) simply imposes the equality of these marginal contributions. To see this, rearrange (13) to

$$\frac{B_{q_1}(q_1^*, q_2^*)}{B_{q_2}(q_1^*, q_2^*)} = \frac{D_{q_1}(q_1^*, q_2^*)}{D_{q_2}(q_1^*, q_2^*)},$$

which says that the marginal rates of substitution between the two qualities have to be identical in the social benefit function and in the marginal revenue function.

\(^5\)They are characterized by the first-order conditions:

$$B_{q_1}(q_1^*, q_2^*) - D_{q_1}(q_1^*, q_2^*)C(q_1^*, q_2^*, e^*) - D(q_1, q_2^*)C_{q_1}(q_1^*, q_2^*, e^*) - G_1'(q_1) = 0,$$

$$B_{q_2}(q_1^*, q_2^*) - D_{q_2}(q_1^*, q_2^*)C(q_1^*, q_2^*, e^*) - D(q_1^*, q_2)C_{q_2}(q_1^*, q_2^*, e^*) - G_2'(q_2) = 0,$$

$$-D(q_1^*, q_2^*)C_{e}(q_1^*, q_2^*, e^*) - H'(e^*) = 0,$$

which have the usual interpretations.
The insurer can still implement the first best if it discloses a quality index, rather than full information about the qualities. Suppose that the provider’s qualities are \( q_1 \) and \( q_2 \). Construct the quality index 
\[
\phi q_1 + (1 - \phi) q_2 \equiv K(\phi),
\]
where \( 0 \leq \phi \leq 1 \). The insurer announces this quality index. When consumers observe \( K(\phi) \), they draw inferences about the unobservable qualities \( q_1 \) and \( q_2 \).

Analogous to Lemma 1, the equilibrium inference must be qualities \( \hat{q}_1 \) and \( \hat{q}_2 \) which solve
\[
\begin{align*}
\max_{q_1, q_2} & \quad T + pD(\hat{q}_1, \hat{q}_2) - D(\hat{q}_1, \hat{q}_2)C(q_1, q_2, e) - G_1(q_1) - G_2(q_2) - H(e) \\
\text{subject to} & \quad \phi q_1 + (1 - \phi) q_2 = \phi \hat{q}_1 + (1 - \phi) \hat{q}_2.
\end{align*}
\]
Any choice of qualities that achieve the quality index will yield the same inference. The provider optimally chooses those qualities that maximize profit, given the quality index. A suitable choice of the index weight \( \phi \) therefore can implement the first-best marginal rate of substitution between the two qualities, as in (14).

The insurer next chooses a prospective price. Given that the provider internalizes the cost, a quality index and a prospective payment are sufficient to implement the first best.

Cost reimbursement with value index can perform exactly the same. Here, the insurer constructs a value index: 
\[
I(\theta_1, \theta_2) = \theta_1 q_1 + \theta_2 q_2 + (1 - \theta_1 - \theta_2) c,
\]
where the weights, \( \theta_1 \) and \( \theta_2 \), are positive and sum to less than 1. Under cost reimbursement, equilibrium qualities and cost effort must minimize the disutility. Any equilibrium \( \hat{q}_1, \hat{q}_2, \) and \( \hat{c} \) solve
\[
\begin{align*}
\max_{q_1, q_2, c} & \quad T + mD(\hat{q}_1, \hat{q}_2) - G_1(q_1) - G_2(q_2) - H(e) \\
\text{subject to} & \quad \theta_1 q_1 + \theta_2 q_2 + (1 - \theta_1 - \theta_2) c = \theta_1 \hat{q}_1 + \theta_2 \hat{q}_2 + (1 - \theta_1 - \theta_2) \hat{c}.
\end{align*}
\]
Using the value-index weights, the insurer controls how the provider trades off between each quality and the cost effort, analogous to Lemma 1. Finally, using the margin the insurer implements the first best, as in Proposition 2.

4.4 Dumping and cream-skimming

The equivalence of prospective payment and cost reimbursement no longer holds when cost is uncertain. We can extend the model for cost heterogeneity. In this case, the provider may dump high-cost consumers, and
cream-skim low-cost consumers. We first address the problem of dumping. Let the unit cost of treating a consumer be randomly distributed on $[C, \bar{C}]$. We use $F$ to denote the cumulative distribution, which is a function of both quality and effort. Therefore, $F(C; q, e)$ is the proportion of consumers who can be treated at unit cost below $C$ when quality and effort are, respectively, $q$ and $e$. We let $F(C; q, e)$ be strictly positive on $[C, \bar{C}]$. Under prospective payment, the provider chooses the first-best quality and effort, and accepts all consumers only if $\bar{C} \leq p$. Otherwise, when $p < \bar{C}$ the provider will dump all those consumers with cost above $p$. The first best is not implementable.

Alternatively, the provider can raise profit by cream-skimming low-cost consumers under prospective payment. For simplicity, suppose there are two types of consumers, $A$ and $B$. For fixed quality and effort, the cost of a type-$B$ consumer is $\beta$ times higher than the cost of a type-$A$ consumer, where $\beta > 1$. Here, cream-skimming refers to the use of different quality levels to discriminate against consumers with different costs. Let $q_A$ and $q_B$ be the respective quality levels the provider chooses for the type-$A$ and the type-$B$ consumers. The provider will cream-skim the type-$A$ consumers by setting $q_B < q_A$. Again, the first best is not implementable. Full analysis of dumping and cream-skimming can be found in Ma (1994).

Here, we emphasize that cost reimbursement has none of the problems of dumping and cream-skimming. Under cost reimbursement, the actual treatment cost of a consumer is fully reimbursed. dumping and cream-skimming are unprofitable. By choosing suitable margin $m$ and weight $\theta$, the insurer continues to implement the optimal quality and cost effort, as in Proposition 2.

5 Conclusion

Prospective payment and cost reimbursement are common payment mechanisms for health care services. In the past thirty years, many theoretical and empirical studies have pointed out the different quality and cost incentives of the two payment systems. In this paper, we have shown how, by optimally choosing the content of public report, an insurer can make the two payment systems implement identical quality and cost incentives. Our results are robust to report misinterpretation, unobservable cost effort, and multiple qualities. Because prospective payment is known to create dumping, cream-skimming, and up-coding incentives, our
result is particularly relevant when patient selection problems are serious.

The main point here can be interpreted as using information as an incentive strategy. Given that health service quality is difficult for consumers to know about, it is incumbent upon insurers and regulators to inform consumers. The usual approach is a sort of “empowering” consumers with as much information as common consumer cognition allows. Here, we question this approach. Information disclosure affects a provider’s incentive to invest in quality and cost effort, and should be considered along with payment mechanisms.

Our analysis is based on a linear value index. Linearity is a restriction, but linear functions are analytically tractable. Linear value or quality indexes are likely better understood by consumers than more complicated schemes. Given that we can implement the first best with a linear index, it is not surprising that nonlinear ones may also succeed. A candidate is a kind of “forcing” index. The insurer fully discloses quality if and only if cost effort is no less than the first-best level; otherwise, the insurer discloses nothing. In effect, the insurer threatens to shut down the market if the provider refuses to choose the first-best cost effort.

The forcing index lacks credibility and robustness. First, it seems incredible that an insurer can commit to such a drastic measure as effectively shutting down the market. Second, the discontinuity in the forcing index is unattractive. If a provider chooses a cost effort slightly lower than the first best, the outcome becomes untenable. Our linear index, however, is robust. A small error in model specification will only lead to a small deviation in the equilibrium.

We have assumed that the insurer can make a lump-sum transfer to the provider. This is consistent with the vast majority of the literature on provider payment design. Two recent papers study optimal provider payment systems when lump-sum transfer is not allowed. Mougeot and Naegelen (2005) show that the first-best quality and cost effort are not attainable without transfer. They then characterize the constrained-optimal prospective price and margin. Miraldo et al. (2011) further characterize the constrained-optimal prospective price list when providers have different cost types. In our model, the first best may not be achieved when transfer is not allowed; a single prospective price or margin cannot handle both distribution and incentive problems. Yet, value-index reporting will continue to induce cost-reduction effort under cost reimbursement.
As the health care market evolves, payment systems have tended to become complicated. Pay for performance is now discussed often in policy and theoretical research; see, for example, works by Eggleston (2005), Kaarboe and Siciliani (2011), and Richardson (2011). The idea in this paper calls for a more fundamental approach. Any reward system must be based on available information. A central issue, as we have shown here, is how the insurer may strategically disclose information. Furthermore, information and financial instruments should be chosen simultaneously to align incentives.
Appendix

Proof of Lemma 1: Suppose that equilibrium quality \( \hat{q} \) and effort \( \hat{e} \) achieve the value index \( I(\theta) = \theta \hat{q} + (1 - \theta) \hat{e} \). Consumers must believe, in equilibrium, that the provider’s quality is \( \hat{q} \), so the demand is \( D(\hat{q}) \). Next, suppose that the provider deviates to any other pair of quality and effort such that the same value index is achieved. That is, suppose the provider deviates to any \( q \) and \( e \) where \( \theta \hat{q} + (1 - \theta) \hat{e} = \theta q + (1 - \theta) e \), then consumers must continue to believe that the quality is \( \hat{q} \). The provider’s profit is \( T + mD(\hat{q}) - G(q) - H(e) \).

By definition of an equilibrium, \( T + mD(\hat{q}) - G(q) - H(e) \leq T + mD(\hat{q}) - G(q) - H(e) \). Because \( G \) and \( H \) are strictly convex, the inequality is strict if and only if \( (q, e) \neq (\hat{q}, \hat{e}) \). Maximizing (7) subject to \( I(\theta) = \theta q + (1 - \theta)e \), we obtain the first-order condition (8).

Proof of Proposition 2: First, by Lemma 1, to ensure that the first-best quality \( q^* \) and effort \( e^* \) can be an equilibrium choice by the provider, the value of the weight must satisfy

\[
\frac{G'(q^*)}{H'(e^*)} = \frac{\theta}{1 - \theta}.
\]

Solving this equation for \( \theta \) yields the value for \( \theta^* \) in the Proposition. For the rest of the proof, \( \theta \) is set at this value.

Second, again from Lemma 1, for any \( \theta \) we use (8) to define implicitly \( e \) as a function of \( q \). (This function also depends on \( \theta \), but now that \( \theta \) is fixed at \( \theta^* \) we omit \( \theta \) from the argument of function.) This yields \( e = e(q) \), with

\[
e'(q) = \frac{1 - \theta G''(q)}{\theta H''(e)} > 0.
\]

For any given \( m \), the provider’s objective can now be regarded as a choice of \( q \) that maximizes

\[
T + mD(q) - G(q) - H(e(q)).
\]

The first-order condition is

\[
mD'(q) = G'(q) + H'(e) \times e'(q).
\]

The right-hand side of (19) is strictly positive for any \( q \). Because \( D' \) is positive, there must exist \( m > 0 \) to satisfy (19) at any \( q \). The value of \( m^* \) in the Proposition is the solution for \( m \) in (19) at \( q = q^* \) and
\( \theta = \theta^* \). The expression for \( m^* \) in the Proposition is obtained after simplification by using (17) and the identity \( \ln f(x) \equiv f'(x)/f(x) \).

Finally, the value of the transfer \( T \) is chosen that \( T + m^* D(q^*) - G(q^*) - H(e^*) = 0 \).
Supplement: multiple qualities and equilibria

In this supplementary note, we illustrate how to implement the first best with multiple qualities.

Prospective payment

From the maximization program (15), we obtain the first-order condition
\[
\frac{D(q_1^*, q_2^*)C_{q_1}(q_1, q_2, e) + G'_1(q_1)}{D(q_1^*, q_2^*)C_{q_2}(q_1, q_2, e) + G'_2(q_2)} = \frac{\phi}{1 - \phi}.
\]

Setting the qualities and effort to \((q_1^*, q_2^*, e^*)\) and rearrange terms, we get the equilibrium weight
\[
\phi^* = \frac{D(q_1^*, q_2^*)C_{q_1}(q_1^*, q_2^*, e^*) + G'_1(q_1^*)}{D(q_1^*, q_2^*)[C_{q_1}(q_1^*, q_2^*, e^*) + C_{q_2}(q_1^*, q_2^*, e^*)] + G'_1(q_1^*) + G'_2(q_2^*)}.
\]

Given \(\phi^*\), the provider’s constrained-maximization problem is
\[
\max_{q_1, q_2} T + pD(q_1, q_2) - D(q_1, q_2)C(q_1, q_2, e) - G_1(q_1) - G_2(q_2) - H(e)
+ \lambda\{\phi^*[D(q_1, q_2)C_{q_2}(q_1, q_2, e) + G'_2(q_2)] - (1 - \phi^*)[D(q_1, q_2)C_{q_1}(q_1, q_2, e) + G'_1(q_1)]\}.
\]

The equilibrium \(p^*\) that implements the first best can be obtained straightforwardly by solving the first-order conditions of the maximization problem. The transfer \(T\) is chosen such that
\[
T + p^*D(q_1^*, q_2^*) - D(q_1^*, q_2^*)C(q_1^*, q_2^*, e^*) - G_1(q_1^*) - G_2(q_2^*) - H(e^*) = 0.
\]

Cost reimbursement

From the maximization program (16), we obtain the first-order conditions
\[
\frac{G'_1(q_1)}{\theta_1} = \frac{G'_2(q_2)}{\theta_2} = \frac{H'(e)}{1 - \theta_1 - \theta_2}.
\]

Setting the qualities and effort to \((q_1^*, q_2^*, e^*)\), the first-order conditions give the equilibrium weights
\[
\theta_1^* = \frac{G'_1(q_1^*)}{G'_1(q_1^*) + G'_2(q_2^*) + H'(e^*)},
\]
\[
\theta_2^* = \frac{G'_2(q_2^*)}{G'_1(q_1^*) + G'_2(q_2^*) + H'(e^*)}.
\]
Given $\theta_1^*$ and $\theta_2^*$, the provider’s constrained-maximization problem is

$$\max_{q_1, q_2} T + mD(q_1, q_2) - G_1(q_1) - G_2(q_2) - H(e)$$

$$+ \mu_1[\theta_2^* G'_1(q_1) - \theta_1^* G'_2(q_2)]$$

$$+ \mu_2[(1 - \theta_1^* - \theta_2^*)G'_2(q_2) - \theta_2^* H'(e)].$$

Again, the equilibrium $m^*$ can be obtained by solving the first-order conditions of the maximization problem. The transfer $T$ is chosen such that $T + m^* D(q_1^*, q_2^*) - G_1(q_1^*) - G_2(q_2^*) - H(e^*) = 0.$
References


Dafny L, Dranove D. Do report cards tell consumers anything they don’t already know? The case of Medicare HMOs. RAND Journal of Economics 2008;39; 790-82.


Richardson S. Integrating pay-for-performance into health care payment systems. Mimeo, Department of Health Care Policy, Harvard University, 2011.
Rochaix L. Information asymmetry and search in the market for physicians’ services. Journal of Health Economics 1989;8; 53-84.

