Incentives in Health Care Payment Systems

Ching-to Albert Ma
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
USA
ma@bu.edu

Henry Y. Mak
Department of Economics
Indiana University-Purdue University Indianapolis
425 University Boulevard
Indianapolis, IN 46202
USA
makh@iupui.edu

May 2018

Keywords: Health Care Payment, Incentives, Prospective Payment, Cost Reimbursement, Selection, Creaming, Skimping, Information Disclosure

Acknowledgement: For their comments and suggestions, we thank Randy Ellis, Jacob Glazer, and Iris Kesternich.

Summary

Health services providers receive payments mostly from private or public insurers rather than patients. Provider incentive problems arise because an insurer misses information about the provider and patients, and has imperfect control over the provider's treatment, quality, and cost decisions. Different provider payment systems, such as prospective payment, capitation, cost reimbursement, fee-for-service, and value-based payment, generate different treatment quality and cost incentives. The important issue is that a payment system implements an efficient quality-cost outcome if and only if it makes the provider internalize the social benefits and costs of services. Thus, the *internalization principle* can be used to evaluate payment systems across different settings.

The most common payment systems are prospective payment, which pays a fixed price for service rendered, and cost reimbursement, which pays according to costs of service rendered. In a setting where the provider chooses health service quality and cost-reduction effort, prospective payment satisfies the internalization principle but cost reimbursement does not. The reason is that prospective payment forces the provider to be responsible for cost, but cost reimbursement relieves the provider of the cost responsibility. Beyond this simple setting, the provider may select patients based on patients' cost heterogeneity. Then neither prospective payment nor cost reimbursement achieves efficient quality and cost incentives. A mixed system which combines prospective payment and cost reimbursement performs better than each of its components alone.

In general, the provider's preferences and available strategies determine if a payment system may achieve internalization. If the provider is altruistic towards patients, prospective payment can be adjusted to accommodate altruism when the provider's degree of altruism is known to the insurer. However, when the degree of altruism is unknown, even a mixed system may fail the internalization principle. Also, the internalization principle fails under prospective payment when the provider can upcode patient diagnoses for more favorable prices. Cost reimbursement attenuates the upcoding incentive. Finally, when the provider can choose many qualities, either prospective payment and cost reimbursement should be combined with the insurer's disclosure on quality and cost information to satisfy the internalization principle.

When good health care quality is interpreted as a good match between patients and treatments, payment design is to promote good matches. The internalization principle now requires the provider to bear benefits and costs of diagnosis effort and treatment choice. A mixed system may deliver efficient matching incentives. Payment systems necessarily interact with other incentive mechanisms such as patients' reactions against the provider's quality choice and other providers' competitive strategies. Payment systems then become part of organizational incentives.

1 Introduction

Incentives play a key role in a health care provider's quality and cost efficiencies. The market generally cannot be relied upon to deliver the required incentives due to health insurance covering much of consumers' health expenses, and missing information. An insurer, or a regulator, must design a payment system to deliver the correct incentives for a provider. Efficiency hinges on the payment system design.

A payment system compensates or rewards a provider based on some observed or verifiable outcomes. These payments can be based on the actual service cost, an episode of care, some information about quality and cost, etc. We will elucidate an *internalization principle*: a payment system achieves efficiency if and only if it manages to get the provider to internalize the social benefit and cost. We will show that the existing literature actually revolves around this principle, either when a payment system succeeds to conform to it, or when it fails.

In the next section, we lay out a simple model, and derive the first best. Then, in Sections 3 and 4 we show how prospective payment satisfies the internalization principle, and how selection may invalidate prospective payment's efficiency claim. Section 5 contains various extensions. Although our presentation uses a theoretical approach, we include some empirical and experimental discussions where appropriate. A few concluding remarks are in Section 6.

2 Model of provider incentives and first best

2.1 Players and technology

We present a canonical model based on Ma (1994). The key "players" are a health care provider and an insurer. We are interested in studying the provider's choices under various incentive schemes designed by the insurer. The provider may be a hospital, a physician, or an allied health professional. Clearly, our abstract provider notion is for simplification, and in specific applications institutional features should be considered. The provider's choices are health care treatments and their qualities, cost reduction efforts, reimbursement coding, and dumping and cream-skimming of patients. The insurer's possible incentive schemes are fee for service or cost reimbursement, prospective payment, and a mixed system. In different subsections, we focus on different subsets of the provider's choices, and begin with quality, and cost effort.

The insurer would like the provider to serve a set of consumers. Let the variable $q \geq 0$ denote care quality. Consumers demand $D(q)$ units of service when the care quality is q, with D being an increasing and concave function. The demand for health services also depends on copayments, deductibles, coinsurance rates, or their combinations. For now, we suppress these demand drivers. Let $B(q)$ denote the social benefit at quality q, where B is increasing and concave. Generally, B is simply consumer surplus, but in specific cases, B may include externality, and such considerations as access and fairness. Again, social surplus B may depend on demand drivers, which are suppressed.

Besides quality, the provider also chooses a cost-reduction effort $e \geq 0$. The unit cost for service at quality q and cost effort e is $C(q, e)$. The function C is strictly increasing in q and strictly decreasing in e, and strictly convex. A higher care quality requires a higher unit cost, but cost effort can reduce it. In addition, the provider incurs a fixed cost or disutility due to quality and cost effort, denoted by $H(q, e)$. The function H is strictly increasing and strictly convex. This basic technology and demand descriptions will be used throughout.

2.2 First best

For a benchmark, let the insurer's objective be social welfare which is benefit less costs,

$$B(q) - D(q)C(q,e) - H(q,e). \tag{1}$$

In the first best, the insurer directly chooses quality and cost effort to maximize social welfare (1). The first-best quality and cost effort (q^*, e^*) are characterized by the two first-order conditions:

$$B'(q^*) = D'(q^*)C(q^*,e^*) + D(q^*)C_q(q^*,e^*) + H_q(q^*,e^*) \tag{2}$$

$$-D(q^*)C_e(q^*,e^*) = H_e(q^*,e^*), \tag{3}$$

where a prime denotes a derivative and subscribes denote partial derivatives. A higher care quality increases social benefit, the left-hand-side term of (2), but also raises demand, unit cost, and disutility, the left-hand-side terms of (2). Condition (2) balances these marginal effects. Next, a higher cost effort reduces unit cost, the left-hand-side term of (3), but raises disutility, the right-hand-side term of (3). Condition (3) balances these marginal effects.

The first best (q^*, e^*) is what the insurer would like to implement.[1] A payment system implements the first best or any other allocation if the system induces the provider to make zero profit and choose the quality and cost effort prescribed by that allocation.

3 Incentives and internalization: prospective payment

In prospective payment, the provider receives a price p per unit of treatment, and a lump-sum transfer T, irrespective of the provider's quality, cost effort, or unit cost. In this section, we let quality and cost effort be the only choices available to the provider. When the provider chooses quality q and cost effort e, its profit is

$$T + pD(q) - D(q)C(q,e) - H(q,e). \tag{4}$$

[1] The insurer's objective can be generalized to maximizing a weighted sum of social net benefit and the provider's profit, with a lower weight on profit. This generalization does not change the implementation problem as long as the insurer can use a lump-sum transfer to extract all profit from the provider. See, for example, Baron and Myerson (1982).

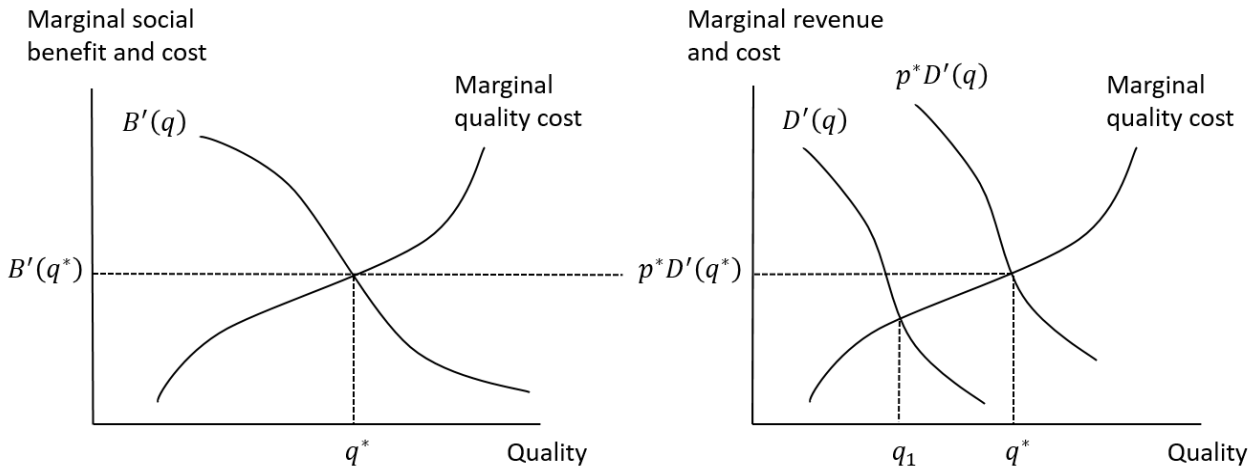

Figure 1: First best and prospective payment

The insurer's objective is to choose the prospective price p and the transfer T to maximize social welfare. The provider must make a nonnegative profit, so the optimal transfer T^* is set to make (4) equal to zero. The value of p influences the provider's choices of q and e to maximize profit. The following proposition is adapted from Ma (1994).

Proposition 1 *By choosing* $p^* = \dfrac{B'(q^*)}{D'(q^*)}$ *and a suitable transfer* T^**, the insurer implements the first best.*

Proof of Proposition 1: The first-order derivative of (4) with respect to q is $pD'(q) - D'(q)C(q,e) - D(q)C_q(q,e) - H_q(q,e)$. For the value of p^* in the Proposition, this first-order derivative vanishes at $(q,e) = (q^*,e^*)$ because we have (2) at $(q,e) = (q^*,e^*)$. Finally, the part of the profit function that involves cost effort e is the same as the part of social welfare that involves e. Hence, at $(q,e) = (q^*,e^*)$, the derivative of the profit function with respect to e vanishes because of (3). We conclude that the profit-maximizing quality and cost effort are first best.□

The intuition is well documented in the literature. Under prospective payment, the provider fully internalizes the total cost $D(q)C(q,e) + H(q,e)$, so its incentive on cost efficiency aligns with the insurer's. Now consider quality incentives. In the first best, the insurer chooses quality q^* to balance the marginal quality cost, $D'(q)C(q,e) + D(q)C_q(q,e) + H_q(q,e)$, and the marginal social quality benefit, $B'(q)$, as depicted in the left panel of Figure 1. Under prospective payment, the provider fully internalizes the marginal

quality cost. The provider's marginal revenue from quality is $pD'(q)$, which is generally different from the marginal social benefit $B'(q)$. Consider an arbitrarily chosen prospective price, say $p = 1$. This is unlikely to align the provider's private benefit and the social benefit. This is shown on the right panel of Figure 1: at $p = 1$, the marginal revenue $D'(q)$ is below $B'(q)$, so the profit-maximizing quality q_1 is also below q^*. However, by setting the prospective price at the value of p^* in Proposition 1, the insurer makes sure that the provider's optimal quality choice internalizes the social benefit at the first-best level. The first best is thereby implemented. Any profit from the prospective payment is taxed away by the lump-sum transfer.

The Proposition is the cornerstone of much of the literature. On the one hand, it clearly elucidates the internalization principle: if a provider can be made to internalize the social interest, it acts in the social interest. Prospective payment is an instrument for the internalization principle to work. On the other hand, the first-best implementation result above focuses only on quality and cost effort. In practice and in general, a provider has other strategic choices, and the insurer has to handle other issues. We now turn to these.

4 Selection and efficiency tradeoff

In this section, we expand the provider's choices to study selection. Unlike many firms that produce physical goods, a health care provider supplies services. Service costs may depend on patient characteristics such as medical history, comorbidity, age, and socioeconomic background. Selection refers to a provider's strategic reaction against heterogenous patient service costs.

4.1 Dumping

For selection, we add diagnostic information to the model, similar to the approaches in Ma (1994) and Ma and Mak (2015) (but see Newhouse (1996) for a synthesis of early literature). We let the provider obtain a patient's diagnostic information, which is the random variable x on $[0, 1]$ with distribution F. Selection is captured by costs varying according to diagnostic information: a patient with diagnosis x has a treatment cost $c(x; q, e)$ at care quality q and cost effort e, for some function c. Furthermore, as a matter of convention, higher values of x mean higher illness severities, so $c(x; q, e)$ is strictly increasing in x, We can interpret the cost function in the last section as the expected service cost $C(q, e) = \int_0^1 c(x; q, e)\mathrm{d}F(x)$ when the provider

serves all patients. We assume that the insurer would like patients with any diagnosis to be treated, so the first best remains unchanged.

Dumping refers to the provider refusing treatment to costly patients. At prospective price p, there may be a diagnosis $\underline{x}$ with $c(\underline{x}; q, e) = p$, so patients with diagnoses higher than $\underline{x}$ will cost more than p; to avoid losses, the provider dumps those with diagnoses $x > \underline{x}$. Under the assumption that the disutility or fixed cost from quality and cost effort is incurred before dumping happens, the provider's profit function becomes

$$\begin{aligned} & T + D(q) \int_0^{\underline{x}} [p - c(x; q, e)] \mathrm{d}F(x) - H(q, e) \\ = \quad & [T + pD(q) - D(q)C(q, e) - H(q, e)] - D(q) \int_{\underline{x}}^1 [p - c(x; q, e)] \mathrm{d}F(x). \end{aligned}$$

Here, at quality q the total demand is $D(q)$, but only $F(\underline{x})$ of those demanding services have costs below p, so the provider makes $p - c(x; q, e)$ for those with diagnosis $x < \underline{x}$. In contrast to the profit without dumping in (4), the provider avoids the loss $D(q) \int_{\underline{x}}^1 [p - c(x; q, e)] \mathrm{d}F(x)$. Prospective payment cannot achieve the first best so it fails the internalization principle.[2]

Due to patient cost heterogeneity, dumping allows the provider to select more profitable cases. The effect on quality, however, cannot be easily determined. The provider does have an incentive for raising quality, attracting more demand, and then rejecting high-cost patients. On the other hand, the provider may reduce quality if quality predominantly attracts high-cost patients.

A number of papers find evidence of dumping under prospective payment, but the magnitude is generally small. The United States introduced the Medicare Prospective Payment System in 1983. Newhouse and Byrne (1988) find that the reduction in length of stay under the System was partly due to a shift of high-cost patients to hospitals exempted from the System.[3] In England, public and private facilities receive the same prospective price for the same procedure. Street et al. (2010) find that both public and private specialized

[2] Dranove (1987) considers a model in which there are both efficient hospitals and inefficient hopsitals, and each hospital can dump patients. He derives the optimal prospective price that balances the social costs of patient allocation inefficiency against dumping.

[3] In the 1990s, some state Medicaid programs started capitated mental health carve-out programs. Ray et al. (2003) find that the Tennessee carve out led to loss of treatment continuity, especially among high-risk patients. Sorbero et al. (2003) find that high-utilization patients with capitated primary care physicians (PCPs) are more likely to switch PCPs than similar patients with fee-for-service PCPs.

treatment centers serve less complex patients than public acute care hospitals.[4]

4.2 Creaming and skimping

We next modify the standard model to consider creaming and skimping. Following Ellis (1998), we define creaming as over-provision of quality to low-cost patients, and skimping as under-provision of quality to high-cost patients. We assume that the insurer wants to provide uniform quality to all consumers, so the first best above remains the same. To focus on quality choices, we suppress dumping here.[5]

At the heart of creaming and skimping is quality discrimination, the provider setting quality according to patients' diagnostic information, say according to the function $q(x)$. Creaming refers to $q(x) > q^*$, and skimping refers to $q(x) < q^*$. At the quality schedule $q(x)$, patients with diagnosis x demand $D(q(x))$ units of service, and their cost is $c(x; q(x), e)$, and the disutility or fixed cost H depends on the profile of qualities $q(x)$. At prospective price p, the provider's profit function is

$$T + \int_0^1 \{D(q(x))[p - c(x; q(x), e)] - H(q(x), e)\} \, \mathrm{d}F(x). \tag{5}$$

By pointwise optimization with respect to $q(x)$, at an interior solution, the profit-maximizing quality at diagnosis x satisfies

$$D'(q(x))[p - c(x; q(x), e)] = D(q(x))c_{q(x)}(x; q(x), e) + H_{q(x)}(q(x), e).$$

The left-hand term is the demand increment from raising quality multiplied by the price-cost margin, so it is quality $q(x)$'s marginal effect on profit. The right-hand term is the marginal quality cost. When x is small so the price-cost margin $p - c(x; q^*, e)$ is high, the provider can profit from raising $q(x)$. This is creaming. Conversely, when x is large so the price-cost margin is low, the provider raises profit by reducing $q(x)$ from q^* to suppress service demand. This is skimping. The provider's profit-maximizing quality schedule seldom results in a constant quality schedule. Prospective payment fails the internalization principle when the provider can cream and skim patients.

[4] Siciliani et al. (2013) further show that treatment centers have shorter length of stay compared to hospitals, after patients' conditions are controlled for. The authors interpret the differences as efficiency gain.

[5] In Ellis (1998), two identical providers choose diagnosis-specific qualities to compete for a fixed number of patients. The equilibrium quality schedules lead to dumping, creaming, and skimping.

The evidence on creaming and skimping is sparse. Frank and Lave (1989) provide one of the first studies based on variation in reimbursement methods across state Medicaid programs in the U.S. They show that the use of prospective payment has increased length of hospital stay for healthier patients (creaming), and decreased length of stay for sicker patients (skimping).

4.3 Cost reimbursement and selection

Under cost reimbursement, the insurer reimburses the provider's actual service cost, $c(x; q, e)$, which is now assumed to be *ex post* verifiable. The disutility or fixed cost is assumed to be either unobserved or noncontractible, so cannot be directly reimbursed. However, the insurer may pay a margin m over the variable cost to cover its fixed cost or disutility. Clearly, when all variable cost $c(x, q, e)$ is reimbursed, the provider has no incentive to incur cost effort, so $e = 0$. Cost reimbursement fails internalization principle, so the first best cannot be implemented.

Selection incentives arise from cost heterogeneity. However, under cost reimbursement, these variations are of no consequence. When the margin m is positive, the provider has no incentive to dump patients. Furthermore, quality discrimination will not be used, again, because cost heterogeneity is of no consequence. The provider's profit function is simply $T + mD(q) - H(q, e)$.

The difference between prospective payment and cost reimbursement is this. Prospective payment lets the provider internalize the cost when selection is absent. However, when cost heterogeneity leads to selection, the provider may practice dumping, creaming, and skimming, so will not internalize all costs. Cost reimbursement results in muted cost incentives, but eliminates selection incentives. The insurer can still use the profit margin m to incentivize quality.[6] However, the quality incentives under cost reimbursement and prospective payment are generally not comparable.

A large empirical literature uses the U.S. Medicare hospital payment reform in 1983 to study the incentive properties of cost reimbursement and prospective payment (for a summary, see Table 7 of Cutler and

[6] Consider the welfare maximization problem. At zero cost effort, the social welfare under cost reimbursement is $B(q) - D(q)C(q, 0) - H(q, 0)$. The insurer maximizes welfare by choosing T and m to implement quality q^C, which is characterized by the first-order condition, $B'(q^C) - D'(q^C)C(q^C, 0) - D(q^C)C_q(q^C, 0) - H_q(q^C, 0) = 0$. Because the marginal social cost of quality is higher without cost effort, the welfare-maximizing quality q^C is below q^*.

Zeckhauser (2000)). The literature finds that there in generally less selection under cost reimbursement, and that the average length of stay under cost reimbursement is about 25% longer. However, the effect of payment system change on average quality is ambiguous.[7] At the physician level, both McGuire (2000) and Legar (2008) review empirical studies that compare doctor behavior under the two payment systems. These studies show that treatment intensity is higher under cost reimbursement.[8, 9]

4.4 Mixed payments: two-part tariffs

Prospective payment and cost reimbursement create different selection and cost-effort incentives. Each payment system may not be completely consistent with the internalization principle. The literature has considered a combination of prospective payment and cost reimbursement, a so-called mixed system, to balance between selection and efficiency.[10] A mixed payment works similarly to a two-part tariff: it consists of a fixed price p and a cost share s, $0 \leq s \leq 1$. For each service the provider supplies at cost $c(x; q, e)$, the provider gets paid the price p and a fraction $1 - s$ of the cost $c(x; q, e)$. At $s = 1$, the mixed system becomes pure prospective payment; at $p \geq s = 0$ it becomes pure cost reimbursement.

In the mixed system, without the dumping option, the provider's profit function is

$$T + \int_0^1 \{D(q(x))[p - sc(x; q(x), e)] - H(q(x), e)\} \, \mathrm{d}F(x).$$

The mixed system will not entirely eliminate creaming and skimping. Cost heterogeneity is attenuated, but

[7] Other than between-system comparison, McClellan (1997) points out that the Medicare Prospective Payment System does reimburse the costs of intensive treatments. Acemoglu and Finkelstein (2008) show how this arrangement affects hospitals' technology choices. In particular, the introduction of Prospective Payment System has led hosptials to reduce labor inputs (which hospitals internalize the full costs), and to adopt new treatment technologies (which are subsidized by the System).

[8] Mougeot and Naegelen (2005) show that cost reimbursement, together with an optimal globle budget cap, can implement the same quality and cost effort as prospective payment. For examples of globle budget sytems in use, see Chen and Fan (2016).

[9] Hodgkin and McGuire (1994) and Cohen and Spector (1996) study how intensity and quality of hospital and nursing home care are affected by both payment systems and rates. Clemens and Gottieb (2014) and Brekke et al. (2017) study how intensity and quality of physician care are affected by fee-for-service payment rates.

[10] See, for example, Sharma (1998) and Ma (1998).

not erased, and an optimal quality schedule generally differs from the first-best quality.[11] The mixed system will not entirely eliminate dumping either. The provider bears the cost $sc(x; q(x), e)$, and the prospective payment is p. Hence whenever the severity is so high that $sc(x; q(x), e) > p$, the patient is unprofitable.

In 1999, the Quebec government in Canada introduced an optional mixed payment system as an alternative to cost reimbursement. The switch corresponds to a higher s in our model. Dumont et al. (2008) find that physicians who self-selected into the mixed payment system reduced their service volume by 6.15% but raised their average time spent per patient by 3.81%. These correspond to increases in both efficiency and quality in our framework. In a controlled laboratory environment where self-selection is muted, Brosig-Koch et al. (2017) show that both medical and non-medical students overprovide services under cost reimbursement, underprovide services under prospective payment, and deviate the least from the patient-optimal quantity under mixed payment.[12]

5 Extensions

5.1 Provider altruism

Whereas the self-interest preference paradigm is standard in economics, the health economics literature has long recognized that health care providers are also interested in their patients' benefits (see, for example, Arrow, 1963; Ellis and McGuire, 1986; and Chalkley and Malcomson, 1998a). What are the implications of provider altruism on incentives and payment mechanisms? Fundamentally, we will show that the internalization principle remains valid.

We return to the basic model in Section 3 where selection is omitted. We let an altruistic provider's

[11] At an interior solution for pointwise optimization, the following first-order conditions hold:

$$\begin{aligned} D'(q(x))[p - sc(x; q(x), e)] - D(q(x))sc_{q(x)}(x; q(x), e) - H_{q(x)}(q(x), e) &= 0 \\ \int_0^1 \{-D(q(x))sc_e(x; q(x), e) - H_e(q(x), e)\}\,\mathrm{d}F(x) &= 0. \end{aligned}$$

These generically yield solutions different from the first best.

[12] Chalkley and Malcomson (2002) simulate the cost saving from changing the Medicare Prospective Payment System into a mixed payment system.

utility be a weighted average of profit and social benefit.[13] The weight on profit is 1 and the weight on social benefit is a parameter α between 0 and 1. If the social benefit is $B(q)$, the provider's altruistic benefit is $\alpha B(q)$. We call α the provider's degree of altruism.[14] The altruistic benefit does not count towards social welfare; neither can it be extracted by the insurer (Hammond, 1987; Milgrom, 1993).

When the value of α is fixed, prospective payment may satisfy the internalization principle. Under prospective payment at price p and transfer T, the provider's objective function changes from (4) to

$$[T + pD(q) - D(q)C(q, e) - H(q, e)] + \alpha B(q), \tag{6}$$

where we have added the altruistic benefit $\alpha B(q)$ to profit. Because altruism does not interact with cost, the provider's cost incentive is aligned with the insurer's, as before. However, the altruistic benefit is now an additional quality incentive. To implement the first best, the insurer simply reduces the prospective price to

$$p^* = (1 - \alpha)\frac{B'(q^*)}{D'(q^*)}, \tag{7}$$

which then satisfies the internalization principle.

According to (7), the higher the provider's degree of altruism, the lower the insurer sets the price. Indeed, intrinsic incentive (altruism) can substitute for extrinsic incentive (profit margin).[15] A lower price means a lower profit margin. The insurer may need to adjust the transfer to ensure that the provider makes a nonnegative profit:

$$T^* = D(q^*)C(q^*, e^*) + H(q^*, e^*) - (1 - \alpha)\frac{B'(q^*)}{D'(q^*)}D(q^*), \tag{8}$$

which is increasing in α.[16]

---

[13] Chalkley and Malcomson (1998a) use a general altruistic benefit function different from social benefit function. Ellis and McGuire (1986) use patient benefit.

[14] In our formulation, the provider's benefit is linked to its action so the altruism is impure (Andreoni, 1990). For our purpose, the terms altruism, intrinisic motivation, and warm glow preference can be used interchangeably.

[15] In the extreme case that $\alpha = 1$, the provider values social benefit as much as the insurer, so the insurer sets $p^* = 0$! This can be interpreted as population-based capitation payment or block contract (Chalkley and Malcomson, 1998a).

[16] Ma (1997) studies a similar implementation problem without lump-sum transfer. He shows that the cost of implementing the first best is weakly decreasing in the degree of altruism. He also considers quality incentive in cost reimbursement.

Internalization relies on the price being tailored to the provider's degree of altruism. When the provider's degree of altruism is variable, its degree of altruism will likely be private information. In this case, the provider has an incentive to lie.[17] In fact, a number of papers in the literature have demonstrated that provider's altruism private information makes the first best impossible to implement. See Jack (2005), Choné and Ma (2011), and Liu and Ma (2013).

Next, we turn to an altruistic provider's selection incentives. For a given p and T, suppose that a provider has a patient with cost c, it earns a profit $p - c(q, e)$. Now the social benefit is $B(q)$, and we assume that the altruism that can be attributed to this patient is the average benefit, $\alpha B(q)/D(q)$. Hence, dumping will be exercised when $p - c(q, e) + \alpha B(q)/D(q) < 0$. Dumping then becomes less likely. Because the altruistic benefit is $\alpha B(q)$, it is increasing in quality. Hence, creaming and skimping will also be attenuated. There is even the potential for the provider to choose excessive quality.

Variations of provider quality, cost effort, and total service can be interpreted as consequences of provider altruism variations. For the same patient population, two providers with different degrees of altruism will not internalize the same social benefits and costs. Even complicated mixed payments may not implement the first best.

We have studied altruism in prospective payment for the sake of simplicity. In a model where the degree of altruism is commonly known, Ellis and McGuire (1986) show that mixed payment outperforms both prospective payment and cost reimbursement. In models where the degree of altruism is not known to the insurer, Jack (2005) derives the optimal separating mixed-payment mechanism, but Choné and Ma (2011) show that pooling contract can be optimal.

Empirically, Kolstad (2013) finds that surgeons' quality response to intrinsic incentives is four times larger than their response to profit incentives. In a laboratory experiment where medical students choose treatment quantities under different payment systems, Godager and Wiesen (2013) find substantial variations

[17] For $\alpha = 0$, the provider is not altruistic at all, so the transfer and price will be those in Proposition 1. For $\alpha = 1$, the provider is like the insurer, so the optimal price is 0. Now a selfish provider has an incentive to mimic the fully altruistic provider: getting the large transfer is better.

in subjects' altruism.[18]

5.2 Upcoding and treatment selection

In Section 4, inefficiency occurs because the self-interested provider is paid the same price for treating patients with different diagnoses. In practice, prospective prices are determined by the Diagnosis Related Groups (DRG) (or Healthcare Resource Groups (HRG) in the U.K.), a system that classifies care episodes according to the patients' diagnoses, treatments, and other characteristics. Each DRG determines a price for the provider. For example, a treatment for an episode of pneumonia without complications is one DRG, and a treatment for pneumonia with complication is another DRG. These two DRGs may carry different prices, say a low price for the one without complication, and a higher for the one with complication.

Upcoding refers to a provider misreporting DRGs, and is possible because audits are seldom strict enough to deter manipulation. In the above example, if the provider misrepresents a case of pneumonia without complication as a case with complication, the provider stands to gain due to the higher price for pneumonia with complication. Upcoding invalidates the internalization principle. The optimal price for quality and cost effort for one illness is typically suboptimal for another illness. As in the case of altruism, upcoding may alleviate some problems with dumping because the provider upcodes to a higher price. The effect of upcoding on quality incentive is ambiguous, but unlikely to be welfare neutral.

Since the US Medicare has adopted the DRG system in the 1980s, many diagnosis groups have been refined according to treatment intensity in order to reduce cost heterogeneity within each DRG. For example, a primary disease diagnosis can been split into two diagnosis groups: one for patients receiving surgery and another for patients receiving medicine. Treatment selection refers to a provider picking more profitable treatments for the same illness. In the above example, when surgery and medicine are both effective, a provider may profit by selecting to treat by surgery because the surgery DRG price is higher than medicine.

Malcomson (2005) derives the insurer's optimal prices under treatment selection. He shows that it is generally optimal to set up DRGs according to treatments. However, the optimal price system accommodates

[18] The incentives of not-for-profit hospitals are beyond the scope of this article. For reviews, see Sloan (2000) and Eggleston et al. (2008).

some inefficient treatment choices in order to economize on total provider payment. Hafsteinsdottir and Siciliani (2010) show that refining DRGs according to treatments may be suboptimal when DRG prices are set according to average treatment costs.[19]

Medicare eliminated patient age in the DRG descriptions in 1988. This generated price changes for 43% of Medicare admissions. Dafny (2005) finds that hospitals responded by upcoding patients to more profitable diagnoses, without changing admission volume, quality, or intensity of treatment. Silverman and Skinner (2004) also show that between 1989 and 1996, the share of the most generous DRGs for pneumonia and respiratory infections rose by 10% to 37% among hospitals with different profit statuses.[20] Gilman (2000) studies the impacts of a DRG refinement by Medicaid in 1994. He finds that the refinement lowers the average severities of both high-price and low-price DRGs, which correspond to treatment selection and dumping, respectively.

5.3 Provider limited capacity and quality incentive

In Proposition 1, the prospective price $p^* = \dfrac{B'(q^*)}{D'(q^*)}$ fulfills the internalization principle for the first best. The implicit assumption is that the provider has the capacity to fulfill higher demand when quality is raised. However, when capacity, say X, is limited, then for some quality q, demand may be larger than capacity, $X < D(q)$. Prospective payment cannot make the provider internalize social benefit because the limited capacity prevents the provider to reap the profit from higher demand.

Chalkley and Malcomson (1998b) show that prospective payment can be enriched to restore the internalization principle. The distinction between consumers' demand and the number of consumers that are served is the key. The optimal mechanism pays the provider a lump sum and a fixed price per unit of service provided, as before. The mechanism then adds a payment based on the number of patients who demand services. This extra payment preserves quality incentive when the provider's capacity is reached. Any quality increase affects demand, so results in extra payment to the provider. The additional incentive

[19] Siciliani (2006) derives the optimal split DRG prices that vary accoding to the proportion of intensive treatment performed by a hospital.

[20] Brunt (2011) studies physician payment upcoding under Medicare Part B.

allows prospective payment to fulfill the internalization principle.

5.4 Multiple qualities and information disclosure

In the basic model, we have considered health care quality generally, so it can be regarded as a single general index for the plethora of qualities. How must a payment system deal with incentives for many quality dimensions such as clinical processes, treatment outcomes, patient experiences, and safety? In this subsection, we first show that prospective payment fails to generate the incentive for the efficient quality profile. We then present a remedy. Information disclosure may complement either prospective payment or cost reimbursement for efficiency.

Ma and Mak (2015) incorporate multiple qualities into the model in Section 2.2. Suppose that the treatment has two qualities, q_A and q_B. The notation for demand, social benefit, unit cost, and disutility naturally become $D(q_A, q_B)$, $B(q_A, q_B)$, $C(q_A, q_B, e)$, and $H(q_A, q_B, e)$. The respective concavity and convexity assumptions are maintained.

The social welfare is now

$$B(q_A, q_B) - D(q_A, q_B)C(q_A, q_B, e) - H(q_A, q_B, e). \tag{9}$$

The first-best qualities and cost effort that maximize social welfare, (q_A^*, q_B^*, e^*), are given by the first-order conditions accordingly. The provider's profit under prospective payment is

$$T + pD(q_A, q_B) - D(q_A, q_B)C(q_A, q_B, e) - H(q_A, q_B, e). \tag{10}$$

The provider continues to internalize cost under prospective payment. However, each quality contributes to profit differently: the first-order derivatives of (10) with respect to q_A and q_B are

$$[p - C(q_A, q_B, e)]D_{q_j}(q_A, q_B) - D(q_A, q_B)C_{q_j}(q_A, q_B, e) - H_{q_j}(q_A, q_B, e), \quad j = A, B,$$

where, again, subscripts denote partial derivatives. The first-order derivative of social welfare (9) with respect to qualities are

$$B_{q_j}(q_A, q_B) - C(q_A, q_B, e)D_{q_j}(q_A, q_B) - D(q_A, q_B)C_{q_j}(q_A, q_B, e) - H_{q_j}(q_A, q_B, e), \quad j = A, B.$$

If the first-best qualities are implemented by a single prospective, we must have

$$p = \frac{B_{q_A}(q_A, q_B)}{D_{q_A}(q_A, q_B)} = \frac{B_{q_B}(q_A, q_B)}{D_{q_B}(q_A, q_B)}. \tag{11}$$

For a given prospective price p, each quality contributes differently to the provider's profit. For a given social benefit function B, each quality also contributes differently to social welfare. The requirement in (11) can be written as $\frac{B_{q_A}(q_A, q_B)}{B_{q_B}(q_A, q_B)} = \frac{D_{q_A}(q_A, q_B)}{D_{q_B}(q_A, q_B)}$, which says that the marginal rates of substitution between qualities must be identical in the social benefit and in consumers' valuation functions. Any divergence will violate (11).

One single prospective price generally cannot align the provider's multiple-quality incentives. Generally, prospective payment cannot satisfy the internalization principle. Ma and Mak (2015) propose that correct quality incentives can be implemented when quality information is made available to consumers through the insurer. In the construction, a quality index $Q = \theta q_A + (1 - \theta) q_B$, where θ and $1 - \theta$ are, respectively, the weights on qualities q_A and q_B, is reported to consumers. The weights can be used to realign the marginal-rate-of-substitution divergence through consumers' response towards the quality index.

Information disclosure can be used together with cost reimbursement. Ma and Mak (2015) present such a system. The insurer discloses to consumers a value index, which is a weighted average of qualities and cost efficiency. The index is $V = \theta_A q_A + \theta_B q_B + (1 - \theta_A - \theta_B)[K - C(q_A, q_B, e)]$, where θ_A and θ_B are weights on qualities and K is a unit-cost benchmark, so $K - C$ measures cost reduction. The value index gives information about qualities, but is also affected by cost effort. The provider's optimal way to achieve a value index, therefore, includes positive cost effort—even though cost is fully reimbursed. Because cost reimbursement also eliminates patient-selection incentive, the optimal combination of value index and cost reimbursement can achieve the first best even when selection is possible.[21]

We have maintained the assumption that health care qualities are not contractible by the insurer. Eggleston (2005) studies the optimal design of provider payment when only one of two qualities, say q_A, is contractible. In her model, there is no demand response, the provider is partially altruistic, and the unit

[21] Glazer et al. (2008) consider the optimal design of a consumer satisfaction rating when a provider can practice quality discrimination.

costs of two qualities are separable. She shows that performance pay on q_A causes the provider to raise q_A and lower q_B, the multi-tasking problem identified in Holmstrom and Milgrom (1991). However, the insurer can partially restore the incentive for q_B by lowering the provider's cost share. This result illustrates the importance of balancing incentives across qualities in provider payment design. The literature has considered other aspects of pay for performance in health care. For a comprehensive survey, we refer readers to the article by Luigi Siciliani in this Encyclopedia.

5.5 Diagnostic information and treatment choice as quality

In the basic model, diagnosis is costless and perfectly accurate. In practice, diagnosis decisions often precede treatment choices. Diagnostic information accuracy is part of the overall quality. Costly diagnosis generates social benefit by matching patient severity with the appropriate treatment intensity. It is, however, difficult to incentivize diagnostic effort. This is especially so when the provider who diagnoses also supplies the service. The provider may prefer a more expensive, more intense treatment, so shirk on diagnosis effort.[22] Of course, the incentive misalignment can be mitigated by having the diagnosis and treatment done by different providers. This, however, may not allow the insurer to exploit economies of scope. Jelovac (2001) derives the optimal provider payment when diagnosis and treatment cannot be separated. Her main result is a version of the internalization principle: the provider internalizes the benefit of diagnosis only if there is a supply-side cost share on treatment expense.

Besides diagnosis, matching patients to appropriate treatments counts as quality care. In addition to being cost ineffective, mismatch can also cause harm to the patient. In Liu and Ma (2013), the first best is defined by matching patient to treatment according to severity and intensity. In Subsection 5.2, we observe that the DRG system generates inefficient treatment selection. Liu and Ma (2013) show that a mixed payment can achieve the first best in this setting. This is because the provider does not fully internalize the patient's treatment benefits. Therefore, the first-best matching requires the provider to internalize some, but not all treatment costs. Pflum (2015) incorporates diagnosis decisions and provider competition into

[22] For example, Afendulis and Kessler (2007) show that in cardiac care, diagnosis by an interventional cardiologist rather than a noninterventional cardiologist results in a 10% higher spending, but not better outcomes.

the matching problem. He shows that a mixed payment is again necessary (but not always sufficient) for efficiency.

5.6 Payment incentives and interactions

We have focused on bilateral insurer-provider interactions. Generally, a provider interacts with others. There are four kinds of such interactions. First, the provider interacts with patients. Our use of a demand as a function of the provider's quality is a simplification. Second, the provider may have to coordinate care with other providers who may have contracted with the same insurer. Third, the provider may have to interact with multiple insurers. Fourth, the provider may compete with other providers in the market place. This last issue on competition in the health market is dealt with in other articles in this Encyclopedia, so we will discuss only the first three.

5.6.1 Provider-patient interaction

In our basic model, the provider chooses treatment quality q, which, in turn, determines demand $D(q)$. But in practice, treatment intensity and other clinical decisions are often made by provider and patient jointly, and patient choices are affected by demand drivers such as consumer cost sharing. The complete study of physician-patient interaction is beyond the scope of this article.[23] However, the main issues still illustrate the internalization principle.

Ellis and McGuire (1990), Ma and McGuire (1997), and Ma and Riordan (2002) contain different models of provider interacting with patients. All papers consider the optimal combinations of provider cost sharing (in the mixed payment system) and consumer cost sharing (such as copayment, deductible, and coinsurance). In the simpler models in Ellis and McGuire (1990) and Ma and Riordan (2002), the altruism utility in Subsection 5.1 is reinterpreted as a reduced form of provider-patient interaction. Hence, the provider's objective function is a weighted average of profit and patient's welfare. In both papers, treatment intensity is jointly determined by the provider and patient, and internalization principle can be extended to accommodate the joint decision.

[23] For example, physician-patient interaction is also affected by information (Dranove, 1988; Johnson and Rehavi, 2016) and malpractice liability (Kessler and McClellan, 1996; Currie and MacLeod, 2008). For a detailed survey, see McGuire (2000).

In both models, the insurer uses combinations of provider and consumer cost sharing to realign the two parties' incentives for efficiency.[24] However, both papers show that when the provider gives little weight to patient welfare, provider cost sharing alone is sufficient for efficiency.[25]

In Ma and McGuire (1997), the interaction is more complex. The patient is actively making a decision about treatment intensity, as a response against the provider's quality decision. Health treatment efficiency relies on both provider-quality and patient-intensity decisions. Moreover, the provider and patient may jointly misreport treatment intensity, which provider payment and consumer copayment are based on. Ma and McGuire (1997) show that both provider payment and patient copayment must be carefully designed to align the interrelated incentives for quality, intensity, and reporting.

5.6.2 Interaction among providers

In Subsection 5.5, we have discussed matching patients with different illness severities to different treatment protocols. Often different providers are responsible for different treatments. The prime example is the different roles played by a general practitioner and a specialist. An insurer may contract with many different providers, so different incentives may be offered to generalists and specialists. Referral is generally the process through which a patient is transferred from one provider to another.[26]

In most countries, referrals involve no monetary transfers between providers—even if they both have contracts with the same insurer. A generalist who has a higher cost share has a stronger incentive to refer (Allard et al., 2011; Iversen and Lurås, 2000). Because the generalist does not internalize the specialist's treatment cost, referrals without monetary transfers fail to achieve efficiency.[27]

[24] Iizuka (2007) shows that physicians who dispense drugs in Japan do trade off their own profits from markup against their patients' out-of-pocket costs.

[25] Neither paper considers selection. Using the Ellis and McGuire (1990) framework, Eggleston (2000) shows that addressing selection incentive requires a lower provider cost sharing and a higher consumer cost sharing.

[26] Although referrals are more often made by physicians, our providers can be clinics and hospitals. Facilities that specialize in different treatments are common.

[27] From 1991 to 1998, some generalists in the UK were paid under fundholding. Each fundholder was given a budget and was responsibe for the cost of its patients' specialty care. Malcomson (2004) shows that a fundholder refers too few patients because it internalizes the full cost but not the full benefit of referrals. Dusheiko et al. (2006) find that the abolition of fundholding in 1998 increased ex-fundholders' admission rates for specialist care by between 3.5 and 5.1%.

How is the internalization principle to be applied to guide incentives? The complex relationship between providers calls for an organizational approach. For efficiency, the set of providers who coordinate care must internalize benefits and costs. In recent years, we have seen such an approach in bundled payments to public and private Accountable Care Organizations (ACO).[28] In ACO programs, quality, cost, and coordination incentives are delegated to a group of providers who form an organization to serve a large group of health plan enrollees (say, at least 5,000).

A risk-adjusted capitation payment is the incentive mechanism for internalization.[29] Under a fixed payment, an ACO internalizes the total cost of care. It then incentivizes appropriate treatment choices and care coordination among providers by such financial and nonfinancial instruments as utilization-review and referral protocols, and linking individual provider's remuneration to quality and cost benchmarks.[30] ACO programs are evolving, but some evidence suggests that they have achieved small cost savings (Song et al., 2012; McWilliams et al., 2016).

Provider integration may be efficiency enhancing or deteriorating. Epstein et al. (2010) provide an example in which an organization improves obstetrical care. They find that high-risk patients in group practices are matched with specialists more often than patients of solo physicians, and the better match improves patients' health outcome. However, provider integration can also increase market power, so mute quality and cost incentives. Baker et al. (2016) find that Medicare patients are more likely to choose a low-quality, high-cost hospital when their physician's practice is owned by the hospital. The joint design of provider payment and organization is an important direction for future research.[31]

[28] For a description of ACO, see, for example, https://innovation.cms.gov/initiatives/ACO/. McClellan et al. (2017) consider ACO-like organizations outside the U.S. Carroll et al. (2017) study the cost and utilization effects of a bundled payment reform in Arkansas.

[29] In Medicare ACOs, some fee-for-service payments are still used for physician services, and cost target overruns may be partially reimbursed. These can be regarded as interim steps to achieve full capitation in the longer term.

[30] Frandsen and Rebitzer (2014) study ACO internal incentive design. ACO shares some similarities with managed care, but managed care plans require enrollees to use in-network providers, and use network privileges as a leverage to lower provider prices. For analysis of managed care incentives, see, for example, Cutler et al. (2000), Frank et al. (2000), Ma and McGuire (2002), Miller (2006).

[31] Further discussions and analyses of provider organization and coordination are in Cebul et al. (2008), Grassi and Ma (2016), Jelovac and Macho-Stadler (2002), and Pauly (1979).

Providers also compete with one another. Mak (2018) shows that a version of the internalization principle holds when heterogeneous providers compete by qualities. When consumers can choose among different providers for services, efficiency requires both providers and consumers to internalize their respective service costs and benefits. Therefore, the insurer has to set provider payment and consumer copayment policies jointly. Competition also requires that the optimal provider payment to adjust according to the intensity of competition and market structure.

5.6.3 Interaction among providers and insurers

We have focused on a single insurer's payment design. Many providers contract with many public and private insurers. In this situation, a provider is a common agent of multiple payers.[32] A single provider's payment may have little effect on a provider's quality and cost incentives, and payers may design payment systems strategically. In the common agency framework, Ma and McGuire (1993) and Glazer and McGuire (2002) show that a public insurer can free-ride on private insurers' quality incentives by offering low prices.[33] Frandsen et al. (2017) show that the reliance on cost reimbursement among private insurers can be a result of coordination failures under common agency. In recent years, the Center for Medicare and Medicaid Services has launched and participated in a number of initiatives to align payment mechanisms (Rajkumar et al., 2014).

6 Conclusion

The health market is complex. As Arrow (1963) has argued, the complexity stems from consumers' unobserved health states. However, insurance is still used to smooth consumers' income fluctuations due to medical expenses. Because consumers pay partial medical expenses, the insurer must set payments to providers. This chapter has laid out issues of payment incentives.

[32] A payer also interacts with multiple agents. Negotiation of price and network formation in the private insurer-provider market are beyond the scope of this article. For a comprehensive review, see Gaynor et al. (2015).

[33] Cutler (1998) finds that in the 1980s, a $1 decrease in Medicare payment leads to a $1 increase in private insurers' prices. This supports cost shifting between public and private insurers. However, Clemens and Gottlieb (2017) show that from 1995 to 2002, a $1 decrease in Medicare payment leads to a $1.16 decrease in private insurers' prices. This is consistent with an private insurer-provider bargaining model in which serving Medicare patients is a provider's outside option.

The broad perspective is the internalization principle. Efficiency obtains when the provider is made to internalize the insurer's concern on quality and cost. We compare the pros and cons of the common prospective payment and cost reimbursement mechanisms. Other complications such as dumping, creaming and skimping call for combinations of prospective payment and cost reimbursement. The internalization principle may also be used to evaluate other systems such as valued-based pricing, and information disclosure. It may also shed light on efficiency properties of new innovations such as consumer accessible information technology, genetic tests and services, and personalized medicine.

Further Reading

Ellis, R. P., Martins, B., & Miller, M.M. (2017). Provider payment methods and incentives. In Quah, S.R. (Eds.), *International Encyclopedia of Public Health*, 2(6), 133-143. Academic Press.

Jegers, M., Kesteloot, K., De Graeve, D., & Gilles, W. (2002). A typology for provider payment systems in health care. *Health Policy*, 60(3), 255-273.

Leger, P. T. (2008) Physician payment mechanisms. In Lu, M., & Jonsson, E. (Eds.), *Financing Health Care: New Ideas for a Changing Society* (149-176). Wiley-VCH.

McGuire, T. G. (2000). Physician agency. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (461-536). North Holland.

Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, 34(3), 1236-1263.

Reference

Acemoglu, D., & Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy*, 116(5), 837-880.

Afendulis, C. C., & Kessler, D. P. (2007). Tradeoffs from integrating diagnosis and treatment in markets for health care. *American Economic Review*, 97(3), 1013-1020.

Allard, M., Jelovac, I., & Leger, P. T. (2011). Treatment and referral decisions under different physician payment mechanisms. *Journal of Health Economics*, 30(5), 880-893.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal*, 100(401), 464-477.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5), 941-973.

Baker, L. C., Bundorf, M. K., & Kessler, D. P. (2016). The effect of hospital/physician integration on

hospital choice. *Journal of Health Economics*, 50, 1-8.

Baron, D. P., & Myerson, R. B. (1982). Regulating a monopolist with unknown costs. *Econometrica*, 50(4), 911-930.

Brekke K. R., Holmas, T. H., Monstad, K., & Straume, O.D. (2017). Do treatment decisions depend on physicians' financial incentives? *Journal of Public Economics*, 155, 74-92.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, 26(2), 243-262.

Brunt, C. S. (2011). CPT fee differentials and visit upcoding under Medicare Part B. *Health Economics*, 20(7), 831-841.

Carroll, C., Chernew, M., Fendrick, A. M., Thompson, J., & Rose, S. (2017). Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas (No. w23926). National Bureau of Economic Research.

Cebul, R., Rebitzer, J., Taylor, L., & Votruba, M. (2008). Organizational fragmentation and care quality in the U.S. healthcare system. *Journal of Economic Perspectives*, 22(4), 93-114.

Chalkley, M., & Malcomson, J. M. (1998a). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17(1), 1-19.

Chalkley, M., & Malcomson, J. M. (1998b). Contracting for health services with unmonitored quality. *Economic Journal*, 108(449), 1093-1110.

Chalkley, M., & Malcomson, J. M. (2002). Cost sharing in health service provision: An empirical assessment of cost savings. *Journal of Public Economics*, 84(2), 219-249.

Chen, B., & Fan, V. Y. (2016). Global budget payment: Proposing the CAP framework. *Inquiry*, 53, 1-4.

Choné, P., & Ma, C. T. A. (2011). Optimal health care contract under physician agency. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 101/102, 229-256.

Clemens, J., & Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, 104(4), 1320-1349.

Clemens, J., & Gottlieb, J. D. (2017). In the shadow of a giant: Medicare's influence on private physician payments. *Journal of Political Economy*, 125(1), 1-39.

Cohen, J. W., & Spector, W. D. (1996). The effect of Medicaid reimbursement on quality of care in nursing homes. *Journal of Health Economics*, 15(1), 23-48.

Currie, J., & MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *Quarterly Journal of Economics*, 123(2), 795-830.

Cutler, D. M. (1998). Cost shifting or cost cutting?: The incidence of reductions in Medicare payments. *Tax Policy and the Economy*, 12, 1-27.

Cutler, D. M., McClellan, M., & Newhouse, J. P. (2000). How does managed care do it? *RAND Journal of Economics*, 31(3), 526-548.

Cutler, D. M., & Zeckhauser, R. J. (2000). The anatomy of health insurance. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (563-643). North Holland.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, 95(5), 1525-1547.

Dranove, D. (1987). Rate-setting by diagnosis related groups and hospital specialization. *RAND Journal of Economics*, 18(3), 417-427.

Dranove, D. (1988). Demand inducement and the physician/patient relationship. *Economic Inquiry*, 26(2), 281-298.

Dumont, E., Fortin, B., Jacquemet, N., & Shearer, B. (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, 27(6), 1436-1450.

Dusheiko, M., Gravelle, H., Jacobs, R., & Smith, P. (2006). The effect of financial incentives on gatekeeping doctors: Evidence from a natural experiment. *Journal of Health Economics*, 25(3), 449-478.

Eggleston, K. (2000). Risk selection and optimal health insurance-provider payment systems. *Journal of Risk and Insurance*, 67(2), 173-196.

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1), 211-223.

Eggleston, K., Shen, Y. C., Lau, J., Schmid, C. H., & Chan, J. (2008). Hospital ownership and quality of care: What explains the different results in the literature? *Health Economics*, 17(12), 1345-1362.

Ellis, R. P. (1998). Creaming, skimping and dumping: Provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17(5), 537-555.

Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5(2), 129-151.

Ellis, R. P., & McGuire, T. G. (1990). Optimal payment systems for health services. *Journal of Health Economics*, 9(4), 375-396.

Epstein, A. J., Ketcham, J. D., & Nicholson, S. (2010). Specialization and matching in professional services firms. *RAND Journal of Economics*, 41(4), 811-834.

Frandsen, B., Powell, M., & Rebitzer, J. B. (2017). Sticking points: Common-agency problems and contracting in the US healthcare system (No. w23177). National Bureau of Economic Research.

Frandsen, B., & Rebitzer, J. B. (2014). Structuring incentives within accountable care organizations. *Journal of Law, Economics, and Organization*, 31(suppl_1), i77-i103.

Frank, R. G., Glazer, J., & McGuire, T. G. (2000). Measuring adverse selection in managed health care. *Journal of Health Economics*, 19(6), 829-854.

Frank, R. G., & Lave, J. R. (1989). A comparison of hospital responses to reimbursement policies for Medicaid psychiatric patients. *Rand Journal of Economics*, 20(4), 588-600.

Gaynor, M., Ho, K., & Town, R. J. (2015). The industrial organization of health-care markets. *Journal of Economic Literature*, 53(2), 235-284.

Gilman, B. H. (2000). Hospital response to DRG refinements: The impact of multiple reimbursement incentives on inpatient length of stay. *Health Economics*, 9(4), 277-294.

Glazer, J., & McGuire, T. G. (2002). Multiple payers, commonality and free-riding in health care: Medicare and private payers. *Journal of Health Economics*, 21(6), 1049-1069.

Glazer, J., McGuire, T. G., Cao, Z., & Zaslavsky, A. (2008). Using global ratings of health plans to improve the quality of health care. *Journal of Health Economics*, 27(5), 1182-1195.

Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32(6), 1105-1116.

Grassi, S., & Ma, C. T. A. (2016). Information acquisition, referral, and organization. *RAND Journal of Economics*, 47(4), 935-960.

Hafsteinsdottir, E. J. G., & Siciliani, L. (2010). DRG prospective payment systems: Refine or not refine? *Health Economics*, 19(10), 1226-1239.

Hammond, P. (1987). Altruism. *The New Palgrave: A Dictionary of Economics* (85-87) Macmillan.

Hodgkin, D., & McGuire, T. G. (1994). Payment levels and hospital response to prospective payment. *Journal of Health Economics*, 13(1), 1-29.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7, 24-52.

Iizuka, T. (2007). Experts' agency problems: Evidence from the prescription drug market in Japan. *RAND Journal of Economics*, 38(3), 844-862.

Iversen, T., & Lurås, H. (2000). The effect of capitation on GPs' referral decisions. *Health Economics*, 9(3), 199-210.

Jack, W. (2005). Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, 24(1), 73-93.

Jelovac, I. (2001). Physicians' payment contracts, treatment decisions and diagnosis accuracy. *Health*

Economics, 10(1), 9-25.

Jelovac, I., & Macho-Stadler, I. (2002). Comparing organizational structures in health services. *Journal of Economic Behavior & Organization*, 49(4), 501-522.

Johnson, E. M., & Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1), 115-141.

Kessler, D., & McClellan, M. (1996). Do doctors practice defensive medicine? *Quarterly Journal of Economics*, 111(2), 353-390.

Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7), 2875-2910.

Leger, P. T. (2008) Physician payment mechanisms. In Lu, M., & Jonsson, E. (Eds.), *Financing Health Care: New Ideas for a Changing Society* (149-176). Wiley-VCH.

Liu, T., & Ma, C. T. A. (2013). Health insurance, treatment plan, and delegation to altruistic physician. *Journal of Economic Behavior & Organization*, 85, 79-96.

Ma, C. T. A. (1994). Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1), 93-112.

Ma, C. T. A. (1997). Cost and quality incentives in health care: Altruistic providers. Boston University, Department of Economics. Published in "Incentius de cost i qualitat en l'assistencia sanitaria. Proveidors altruistes," in *La Contractacio de Serveis Sanitaris*, 65-80 and "Incentivos de coste y calidad en la assistencia sanitaria. Proveedores altruistas" in *Contractacio i complra de serveis sanitarias.* L'experiencia comparada i el cas de Catalunya, G. Lopez-Casasnovas (editor), Servei Catala de la Salut de la Generalitat de Catalunya, Barcelona, 1998.

Ma, C. T. A., (1998). Health-care payment systems: Cost and quality incentives—Reply. *Journal of Economics & Management Strategy*, 7(1), 139-142.

Ma, C. T. A., & McGuire, T. G. (1993). Paying for joint costs in health care. *Journal of Economics & Management Strategy*, 2(1), 71-95.

Ma, C. T. A., & McGuire, T. G. (1997). Optimal health insurance and provider payment. *American Economic Review*, 87(4), 685-704.

Ma, C. T. A., & McGuire, T. G. (2002). Network incentives in managed health care. *Journal of Economics & Management Strategy*, 11(1), 1-35.

Ma, C. T. A., & Mak, H. Y. (2015). Information disclosure and the equivalence of prospective payment and cost reimbursement. *Journal of Economic Behavior & Organization*, 117, 439-452.

Ma, C. T. A., & Riordan, M. H. (2002). Health insurance, moral hazard, and managed care. *Journal of Economics & Management Strategy*, 11(1), 81-107.

Malcomson, J. (2004). Health service gatekeepers. *RAND Journal of Economics*, 35(2), 401-421.

Malcomson, J. (2005). Supplier discretion over provision: Theory and an application to medical care. *RAND Journal of Economics*, 36(2), 412-432.

Mak, H. Y. (2018). Managing imperfect competition by pay for performance and referencing pricing. *Journal of Health Economics*, 57, 131-146.

McClellan, M. (1997). Hospital reimbursement incentives: An empirical analysis. *Journal of Economics & Management Strategy*, 6(1), 91-128.

McClellan, M., Udayakumar, K., Thoumi, A., Gonzalez-Smith, J., Kadakia, K., Kurek, N., Abdulmalik, M., & Darzi, A. W. (2017). Improving care and lowering costs: Evidence and lessons from a global analysis of accountable care reforms. *Health Affairs*, 36 (11), 1920-1927.

McGuire, T. G. (2000). Physician agency. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (461-536). North Holland.

McWilliams, J. M., Hatfield, L. A., Chernew, M. E., Landon, B. E., & Schwartz, A. L. (2016). Early performance of accountable care organizations in Medicare. *New England Journal of Medicine*, 374(24), 2357-2366.

Milgrom, P. (1993). Is sympathy an economic value? In Hausman, J. A. (Ed.), *Philosophy, economics,*

and the contingent valuation method. Contingent valuation: A critical Assessment (417-441). North Holland.

Miller, N. H. (2006). Insurer-provider integration, credible commitment, and managed-care backlash. *Journal of Health Economics*, 25(5), 861-876.

Mougeot, M., & Naegelen, F. (2005). Hospital price regulation and expenditure cap policy. *Journal of Health Economics*, 24(1), 55-72.

Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, 34(3), 1236-1263.

Newhouse, J. P., & Byrne, D. J. (1988). Did Medicare's prospective payment system cause length of stay to fall? *Journal of Health Economics*, 7(4), 413-416.

Pauly, M. V. (1979). The ethics and economics of kickbacks and fee splitting. *Bell Journal of Economics*, 10(1), 344-352.

Pflum, K. E. (2015). Physician incentives and treatment choice. *Journal of Economics & Management Strategy*, 24(4), 712-751.

Rajkumar, R., Conway, P. H., & Tavenner, M. (2014). CMS—Engaging multiple payers in payment reform. *Journal of the American Medical Association*, 311(19), 1967-1968.

Ray, W. A., Daugherty, J. R., & Meador, K. G. (2003). Effect of a mental health "carve-out" program on the continuity of antipsychotic therapy. *New England Journal of Medicine*, 348(19), 1885-1894.

Sharma, R. L. (1998) Health-care payment systems: Cost and quality incentives—comment. *Journal of Economics & Management Strategy*, 7(1), 127-137.

Siciliani, L. (2006). Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics*, 25(3), 479-499.

Siciliani, L., Sivey, P., & Street, A. (2013). Differences in length of stay for hip replacement between public hospitals, specialised treatment centres and private providers: Selection or efficiency? *Health Economics*, 22(2), 234-242.

Silverman, E., & Skinner, J. (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics*, 23(2), 369-389.

Sloan, F. A. (2000). Not-for-profit ownership and hospital behavior. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (1141-1174). North Holland.

Song, Z., Safran, D. G., Landon, B. E., Landrum, M. B., He, Y., Mechanic, R. E., Day, M.P. & Chernew, M. E. (2012). The 'Alternative quality contract,' based on a global budget, lowered medical spending and improved quality. *Health Affairs*, 31(8), 1885-1894.

Sorbero, M. E., Dick, A. W., Zwanziger, J., Mukamel, D., & Weyl, N. (2003). The effect of capitation on switching primary care physicians. *Health Services Research*, 38(1 Pt 1), 191–209.

Street, A., Sivey, P., Mason, A., Miraldo, M., & Siciliani, L. (2010). Are English treatment centres treating less complex patients? *Health Policy*, 94(2), 150-157.