

# Managing Imperfect Competition by Pay for Performance and Reference Pricing

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

November 2017

## Abstract

I study a managed health service market where differentiated providers compete for consumers by choosing multiple service qualities, and where copayments that consumers pay and payments that providers receive for services are set by a payer. The optimal regulation scheme is two-sided. On the demand side, it justifies and clarifies value-based reference pricing. On the supply side, it prescribes pay for performance when consumers misperceive service benefits or providers have intrinsic quality incentives. The optimal bonuses are expressed in terms of demand elasticities, service technology, and provider characteristics. However, pay for performance may not outperform prospective payment when consumers are rational and providers are profit maximizing, or when one of the service qualities is not contractible.

**Keywords:** managed competition, multitasking, prospective payment, pay for performance, reference pricing, implementation

**Acknowledgement:** I am grateful to editor Luigi Siciliani, two anonymous referees, Subir Chakrabarti, Ching-to Albert Ma, Xavier Martínez-Giralt, and participants at various seminars and conferences for their helpful comments and suggestions. All errors are my own.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# 1 Introduction

This paper studies a model of managed provider competition with the following characteristics. Consumers obtain health services from competing providers. The copayments that consumers pay, and the payments that providers receive for services, are set by a payer. Each provider chooses its service qualities based on the payer's payment formula and demand response. Consumers, in turn, pick providers based on their service qualities and the copayment levels. Health service markets that fit the description range from the predominantly public system in the United Kingdom, to the mostly private system in the United States.

An important characteristic of health service markets described above is the coexistence of various imperfect incentives. First, providers face competition. But the competition is imperfect due to factors such as distance and unsophisticated consumers. Second, providers may have intrinsic incentives to supply quality services. But their private incentives prevent them from acting as perfect agents for consumers. Third, payers choose both copayments and provider payments to influence demand and supply. However, risk-spreading and other concerns often constrain consumer cost shares to be significantly below service costs. On the other hand, provider payment systems are often based on quantity or imperfect quality measures. These multifaceted incentive problems are further complicated by the heterogeneity of provider productivity, and the multidimensionality of health service qualities that range from clinical process, treatment outcome, patient experience, to safety.

A large literature, reviewed below, studies these incentive problems in a piecemeal manner. The goal of this paper, however, is to analyze the interactions among these incentive issues in a unified model. To model imperfect competition, I consider two providers with different production costs located at the endpoints of a Hotelling line. They compete for consumers by offering services with two endogenous quality dimensions. Consumers have horizontal preferences, and they may overreact or underreact to qualities because of misperceptions. On the other hand, providers can receive non-pecuniary quality benefits, and they may face profit-distribution constraints as not-for-profit organizations. I allow the payer to set a copayment rate for each provider's service, and consider two provider payment mechanisms. Under *prospective payment system*, the payer pays a fixed price for each consumer that a provider serves. Under *pay for performance*, the payer also pays bonuses according to service qualities. However, the payer may not measure one of the two quality dimensions.

I focus on how the payer should coordinate its incentive instruments with market force and providers' intrinsic quality incentives. The payer's goal is to achieve the social optimum, which is defined by provider-specific service quality levels and consumer allocation between providers. The systematic approach reveals the conditions under which a suitable combination of imperfect instruments is sufficient for implementing the optimal outcome. It also illustrates how each of

the payer's instruments must be designed as an integrated part of the incentive system. Many properties of the optimal instruments are new to the insurance benefit design, provider payment mechanism, multi-tasking, and intrinsic motivation literature. I now overview the main results.

*The design of insurance benefit.* In the optimal insurance literature where the supply side is active, a consumer interacts with a given provider to determine treatment quantity, and consumer cost share is used to influence the quantity choice in order to achieve production efficiency (Ellis and McGuire, 1990; Ma and McGuire, 1997). But under imperfect provider competition, each consumer picks a provider for service, and the resulting market shares determine the providers' marginal profits from quality investments. Therefore, consumer cost shares affect both market shares and qualities, which correspond to allocative and production efficiencies, respectively.

I show that both types of efficiencies require consumer cost-sharing to i) reflect the incremental marginal cost when consumers pick a more costly provider for service, and ii) steer misperceived consumers from overvalued service, and to undervalued service. The first requirement justifies the use of differential copayments such as reference pricing and preferred provider network. Moreover, the optimal copayments for using high-cost providers can be below marginal costs and the optimal copayments for using low-cost providers can be as low as zero. The second requirement supports the use of value-based benefit design. However, it clarifies that value-based copayment should be based on misperceived service value rather than value itself.<sup>1, 2</sup>

*The design and limitation of prospective payment system.* On the supply side of health care markets, prospective payment is one of the most common payment mechanisms. A defining characteristic of prospective payment is that it pays a provider irrespective of its service cost and quality levels. It is also well understood that if providers are local monopolies with exogenous quality, production efficiency can be achieved by paying each provider the average cost of comparable providers (Shleifer, 1985). My model allows me to derive the optimal prospective prices when providers compete by choosing qualities. I show that the optimal price of a provider's service depends not only on its competitor's marginal cost, but also on the quality difference at the social optimum (which is fixed) and the intensity of competition. This implies, for example, that the cost-based U.S. Medicare Prospective Payment mechanism systematically underpays high-quality hospitals, and overpays hospitals in more competitive markets.

---

<sup>1</sup>For example, the California Public Employees' Retirement System adopted reference pricing in 2011 (White and Eguchi, 2014). In 2016, 48% of employees who were covered by employer-provided health insurance in the U.S. were enrolled in preferred provider networks (Kaiser, 2016). Thomson et al. (2013) review the international experience of value-based insurance designs.

<sup>2</sup>Baicker et al. (2015) characterize the optimal copayment for insurees who can make mistakes in a model where the supply side is passive. They focus on the interactions of *ex post* moral hazard and consumer mistakes (or behavioral hazard), whereas this paper focuses on the interactions of consumers, providers, and payer.

I also find that prospective payment can achieve the social optimum only under some restrictive conditions such as perfect consumer rationality. In particular, when providers compete by choosing multiple qualities, any degree of consumer misperceptions can distort a provider away from the optimal quality mix, and render the social optimum not attainable by any prospective prices.

*The design of pay for performance, with and without unmeasurable quality.* In recent years, many payers have adopted pay for performance to incentivize health service qualities. However, a review of 34 programs in 14 OECD countries show that performance pay has only generated “mixed, modest and short-lived improvements” (Milstein and Schreyoegg, 2016). On the other hand, theoretical works on pay for performance are almost exclusively grounded on single-agent set up, and hence, they provide little guidance on the optimality of performance pay under imperfect competition.<sup>3</sup> In my model, I first consider the design of performance pay when both service qualities are contractible. I find that the optimal bonus for each quality does depend on the interaction between quality elasticity of demand and service technology. In particular, the optimal bonus rate for a low-marginal-cost service (such as imaging service) is decreasing in demand elasticity, whereas the optimal bonus rate for a high-marginal-cost service (such as surgery) is increasing in demand elasticity.

I then let one of the two qualities be unmeasurable by the payer, which is one of the primary arguments against using performance pay (see, for example, Holmstrom and Milgrom’s (1991) analysis of multitasking). To fix idea, let the two qualities be patient experience and treatment outcome, and the latter be not contractible.<sup>4</sup> I show that a quality bonus on patient experience is optimal if and only if consumers’ preference for treatment outcome is too strong. In particular, substitutability between qualities, an important determinant of the suitability of performance pay in bilateral models, is irrelevant here. Both sets of results, with and without the measurement issue, illustrate the importance of market force in bonus design problems.

*Intrinsic and extrinsic incentives.* The economic literature has long recognized that intrinsic quality incentives, in the form of non-pecuniary quality benefit and organizational profit-distribution constraint, can act as remedies for weak explicit quality incentives (Hansmann, 1996; Glaeser and Shleifer, 2001; Besley and Ghatak, 2005). I instead focus on how payment systems should adapt to providers’ intrinsic incentives in imperfect competitive markets where qualities are multidimensional. Consistent with the literature, I find that intrinsic incentives can reduce the levels of prices and performance bonuses. However, intrinsic incentives also generate paternalis-

---

<sup>3</sup>See, for example, Eggleston (2005). The only two exceptions that I am aware of are Acemoglu et al. (2008), who show that governments have comparative advantage over markets and firms in offering low-powered incentives; and Benabou and Tirole (2016), who consider the impact of labor market competition on performance-pay structure.

<sup>4</sup>When the U.S. Medicare Hospital Value-Based Purchasing Program was launched in 2012, it measured patient experience (and clinical process and safety) but not treatment outcome (VanLare and Conway, 2012). Outcome measure was introduced in 2013.

tic quality preferences and induce providers to deviate from the optimal quality mix. Therefore, the social optimum can be attained only if the payer uses pay for performance to counteract the provider paternalism. In other words, intrinsic and extrinsic incentives are complements rather than substitutes in a multi-quality incentive system.

*Equity concern and provider payment.* In the main model, the social optimum is based on economic efficiency. In an extension, I allow the payer to have a strong equity concern, so that it seeks to implement uniform qualities and sets uniform copayments despite asymmetric provider costs. I show that prospective payment fails to implement the payer’s targeted allocation as if consumers are misperceived or providers are intrinsically motivated. However, the principles for setting performance pay continue to apply after small adjustments are made to account for the change in qualities.

## 1.1 Literature

In addition to the areas described above, this paper also contributes to the literature on managed competition in healthcare (Enthoven, 1993). I use the Hotelling framework to model imperfect competition. This is based on Ma and Burgess (1993). In their main extensive form, two symmetric providers choose unidimensional qualities and then prices, and the providers maximize profits by choosing socially inefficient quality levels. However, a regulator can restore production efficiency by regulating prices and making lump-sum transfers. This framework has been adopted to study many regulation and payment design problems in healthcare markets.

Brekke et al. (2006) allow two providers to choose both unidimensional qualities and locations on the Hotelling line. They find a necessary trade-off between vertical quality and horizontal differentiation when the regulator sets fixed prices before the providers pick qualities and locations. Bardey et al. (2012) show that this trade-off can be mitigated by a mixed-payment scheme. My paper abstracts from location choices, and focuses on providers that offer services with multidimensional qualities.<sup>5</sup> I characterize the conditions under which fixed prospective prices are sufficient for achieving the social optimum.

Wolinsky (1997) compares managed competition, where providers use unidimensional qualities to compete under fixed prices, to local monopolies, where the regulator eliminates competition by dividing the market administratively. He allows providers to have different cost functions as in this paper. He shows that under complete information, local monopolies can achieve the social optimum but managed competition cannot. However, when the regulator is uninformed about the cost functions, managed competition is more effective in extracting information rents, and hence,

---

<sup>5</sup>I also abstract from entry, exit, and merger of providers. Katz (2013) shows that entry can lower market quality. Brekke et al. (2017) show that hospital merger may increase or decrease market quality.

the ranking of the two regimes is ambiguous.<sup>6</sup> I show that managed competition can achieve the social optimum when provider payments are separated from consumer cost shares.<sup>7</sup>

Using a differential-game approach, Brekke et al. (2012a) study quality dynamics under managed competition. They use information lag to provide a foundation for consumers' quality misperceptions, and characterize the dynamic price regulation that implements the optimal quality path. Brekke et al. (2012b) consider competition between symmetric non-profit providers in a static model. They show that when prices are regulated, profit-distribution constraints can only induce very altruistic providers to raise quality. My model incorporates both consumer misperception and provider intrinsic incentives. I show that they both favor pay for performance over prospective payment system.

Other than the mixed effects on qualities, the empirical literature also finds that some providers respond to pay for performance strategically (see, for example, Gravelle et al., 2010; Mullen et al., 2010). Eggleston (2005) shows that mixed payment can mitigate strategic response to performance pay in a moral hazard framework. Wu et al. (forthcoming) add hidden provider productivity to the framework. They characterize a menu of optimal screening contracts in which the supply-side cost share and performance bonus are, respectively, decreasing and increasing in the provider's reported productivity. My paper incorporates demand response into the payer's contract design problem. In particular, I show that quality elasticity of demand is an important determinant of the optimal bonus rate. Recently, Gaynor et al. (2016) and Moscelli et al. (2016) estimate the quality elasticities of demands for heart surgery and hip replacement, respectively.

The structure of the paper is as follows. Section 2 sets up a benchmark model and shows the source of inefficiency under unmanaged provider competition. Section 3 characterizes the optimal copayments, prospective payment system, and pay for performance in the benchmark model. Section 4 incorporates consumer misperceptions, provider intrinsic quality incentives, unmeasurable quality, and alternative market structures into the benchmark model. Section 5 introduces equity concern to both the benchmark and the expanded models. Section 6 draws some concluding remarks. Appendix A collects the proofs of all lemmas and propositions. Appendix B contains the proofs of some auxiliary results.

---

<sup>6</sup>Beitia (2003) compares managed competition to a monopoly that serves the entire market. She shows that information asymmetry distorts the less-efficient firm's market share downward. This distortion can change the optimal market structure from managed competition to monopoly. See also McGuire and Riordan (1995).

<sup>7</sup>Brekke et al. (2011) analyze reference pricing in a vertical differentiation model where the supply side is not active. They also show that reference pricing reduces both brand-name and generic drug prices in Norway.

## 2 Model

### 2.1 Asymmetric providers and differentiated consumers

There are two providers,  $A$  and  $B$ , that supply health services to consumers. The service of provider  $i$ ,  $i = A, B$ , has two quality dimensions,  $q^i$  and  $r^i$ , both chosen by the provider. A consumer's benefit from receiving service with qualities  $(q^i, r^i)$  is  $v_q q^i + v_r r^i$ , where all parameters are strictly positive.<sup>8</sup> In this and the next section, I assume that  $v_q q^i + v_r r^i$  is both the true and perceived benefits from qualities  $(q^i, r^i)$ . I will relax this assumption in Section 4.

In addition to service benefits, each consumer is also described by a proclivity parameter  $x$  towards the two providers *à la* Hotelling. Consumers are uniformly distributed on the  $[0, 1]$  interval, whereas providers  $A$  and  $B$  are located at points 0 and 1, respectively. A consumer at location  $x$  incurs a cost  $\tau x$  to use provider  $A$ , and a cost  $\tau(1 - x)$  to use provider  $B$ . The parameter  $x$  captures all horizontal differentiation factors, such as travel distances and auxiliary services, that affect consumer payoffs. Consumers have different proclivities towards these factors, and  $\tau$  captures the relative importance of  $x$  in consumer choices.

Every consumer has to pay a nonnegative and identical cost share  $\Lambda^i$  to use provider  $i$ . I will expand on the determinations of consumer cost shares in Subsections 2.3 and 3.1. When using providers  $A$  and  $B$ , consumer  $x$ 's payoffs are  $v_q q^A + v_r r^A - \Lambda^A - \tau x$  and  $v_q q^B + v_r r^B - \Lambda^B - \tau(1 - x)$ , respectively. The demand for provider  $i$ 's service is  $D^i$ . Except in Subsection 4.4, I will maintain the assumption that each consumer obtains service from one of the providers, so that the market is covered, and hence,  $D^A + D^B = 1$ .

The total cost of provider  $i$  is determined by a provider-specific, twice-differentiable function  $C^i(D^i, q^i, r^i)$ . The total cost is strictly increasing in demand and each quality ( $C_D^i > 0$ ,  $C_q^i > 0$ ,  $C_r^i > 0$ ), strictly positive, and strictly convex in qualities. In this and the next section, providers can be interpreted as healthcare professionals such as physicians or institutions such as hospitals, and the objective of each provider is to maximize its own profit. I will introduce motivation benefits and profit-distribution constraints in Section 4.

### 2.2 First best

I let social welfare be the sum of consumer payoffs less the sum of production costs. In the first best, a social planner chooses each provider's market share  $D^i$  and qualities  $(q^i, r^i)$  to maximize social welfare. I take the duopoly market structure as given and assume that both providers are

---

<sup>8</sup>All the results in this paper continue to hold if consumer benefit is  $V(q^i, r^i)$ , where  $V$  is twice differentiable, strictly increasing, and concave.

active. The social welfare is

$$\int_0^{D^A} [v_q q^A + v_r r^A - \tau x] dx - C^A(D^A, q^A, r^A) + \int_{1-D^A}^1 [v_q q^B + v_r r^B - \tau(1-x)] dx - C^B(D^B, q^B, r^B). \quad (1)$$

The following lemma characterizes the first best.

**Lemma 1** *The first-best market share  $D_*^i$  and qualities  $q_*^i, r_*^i$  of provider  $i, i = A, B, j \neq i$  satisfy the following:*

$$\frac{1}{2} + \left[ \frac{v_q(q_*^i - q_*^j) + v_r(r_*^i - r_*^j) - \{C_D^i(D_*^i, q_*^i, r_*^i) - C_D^j(D_*^j, q_*^j, r_*^j)\}}{2\tau} \right] = D_*^i \quad (2)$$

$$D_*^i v_q = C_q^i(D_*^i, q_*^i, r_*^i) \quad (3)$$

$$D_*^i v_r = C_r^i(D_*^i, q_*^i, r_*^i). \quad (4)$$

Expression (2), the first-best market share of provider  $i$ , ensures that the optimal marginal consumer at  $x = D_*^A$  receives the same marginal social welfare from either provider. When the two providers are symmetric,  $D_*^A$  must be 1/2. Hence, the squared-bracket term in (2) summarizes the effect of service-cost asymmetry on the first-best consumer allocation. The next two equations show that raising a provider's quality increases its consumers' payoffs, but also the service cost. The first-best qualities of each provider  $i$  balance these marginal effects.<sup>9</sup>

### 2.3 Unmanaged provider competition

Before I study managed competition, I first consider an unmanaged health service market in which consumers are covered by an indemnity plan. The payer sets a uniformed coinsurance rate  $\phi$ ,  $0 < \phi \leq 1$ . Each provider  $i$  then chooses price and qualities simultaneously to maximize its profit. I use Nash equilibrium as the solution concept. The maximization program of provider  $i$  is

$$\max_{\Lambda^i, q^i, r^i} \left[ \frac{1}{2} + \frac{v_q(q^i - q^j) + v_r(r^i - r^j) - \{\Lambda^i - \Lambda^j\}}{2\tau} \right] \left( \frac{\Lambda^i}{\phi} \right) - C^i(D^i, q^i, r^i), \quad (5)$$

where the squared-bracket term is provider  $i$ 's market demand  $D^i$ , and where  $\Lambda^i/\phi$  is provider  $i$ 's per-unit revenue when it sets consumers' coinsurance payment to  $\Lambda^i$ .<sup>10</sup> If the Nash equilibrium outcome coincides with the first best, managing provider competition cannot improve social welfare. A welfare-maximizing payer's role is confined to choosing a coinsurance rate, and lump-sum transfers

<sup>9</sup>It is not possible to pin down the comparative statics of  $\tau$  in the first best without additional assumptions on the model's primitives. See Appendix B.

<sup>10</sup>In Appendix B, I provide sufficient conditions for (5) to be quasi-concave and each provider  $i$ 's best response to be a contraction map. These conditions guarantee the existence and uniqueness of a pure-strategy Nash equilibrium in the unmanaged market game.



to and from consumers and providers. However, the following lemma, which extends Proposition 1 in Wolinsky (1997, p.827) to my environment, shows that unmanaged provider competition is unlikely to achieve the first best.

**Lemma 2** *Under unmanaged provider competition where provider  $i$ 's maximization program is given by (5), the Nash equilibrium  $(\Lambda_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i)$  can achieve the first best  $(D_{*}^i, q_{*}^i, r_{*}^i)$  only if  $D_{*}^i = 1/2$  and  $\phi = 1$ .*

In equilibrium, provider  $i$  sets its consumers' coinsurance payment equal to  $\Lambda_{\dagger}^i = 2\tau D_{\dagger}^i + \phi C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i)$ , and receives  $\Lambda_{\dagger}^i/\phi = 2\tau D_{\dagger}^i/\phi + C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i)$  from serving a consumer. Because  $C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i)$  is the cost of serving the marginal consumer,  $2\tau D_{\dagger}^i/\phi$  can be interpreted as the (marginal) markup. Substituting  $\Lambda_{\dagger}^i$  into the other first-order conditions (in Appendix A) yields the equilibrium values of  $D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i$ , which are characterized by

$$\frac{1}{2} + \left[ \frac{v_q(q_{\dagger}^i - q_{\dagger}^j) + v_r(r_{\dagger}^i - r_{\dagger}^j) - \phi\{C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i) - C_D^j(D_{\dagger}^j, q_{\dagger}^j, r_{\dagger}^j)\}}{6\tau} \right] = D_{\dagger}^i \quad (6)$$

$$D_{\dagger}^i v_q = \phi C_q^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i) \quad (7)$$

$$D_{\dagger}^i v_r = \phi C_r^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i). \quad (8)$$

The differences between these equations and those in Lemma 1 illustrate the source inefficiencies under unmanaged provider competition.

In the first best, the social planner fully internalizes the social costs of consumer allocation and qualities. However, consumer choices under unmanaged competition are driven by the difference in coinsurance payments,  $\Lambda_{\dagger}^i - \Lambda_{\dagger}^j$ . Therefore, consumers are less responsive to difference in costs. This, in turn, induces providers to over invest in qualities to attract consumers. These effects are captured by the parameter  $\phi$  in conditions (6) to (8). Moreover, providers under unmanaged imperfect competition use their market power to price their services above marginal costs. The makeups  $2\tau D_{\dagger}^i/\phi$  further distort market shares and qualities by turning  $2\tau$  in (2) into  $6\tau$  in (6).<sup>11</sup>

The quality distortions due to coinsurance rate below one are consistent with the general equilibrium effects of the introduction of Medicare in 1965 (Finkelstein, 2007). Moreover, Cooper et al. (2015) show that prices in the U.S. private hospital markets are positively associated with hospital market power. In fact, the inefficiencies identified in Lemma 2 generally arise, unless the two providers are symmetric and the coinsurance rate  $\phi$  is set equal to one. These two assumptions are frequently made in the imperfect competition literature. However, symmetry eliminates the squared-bracket term in (6) and generates allocation efficiency trivially. Moreover, (7) and (8) show that when the coinsurance rate is one, the providers' overinvestment incentives are muted.

---

<sup>11</sup>The source of distortion here is different from the one in Spence (1975) because the marginal consumer's quality valuation is representative when consumers and firms are horizontally differentiated.

Starting with the next section, I consider managed competition where the payer actively chooses prices. I will maintain the assumption that the providers are asymmetric.

### 3 Managed competition

I allow the payer to separately set cost shares that consumers pay, and payments that providers receive for services. I continue to use  $\Lambda^i$  to denote the cost share that a consumer pays to use provider  $i$ 's service. But from now on, I interpret  $\Lambda^i$  as the copayment for provider  $i$ 's service. I use the accounting convention that consumer cost share  $\Lambda^i$  is received by the payer rather than the provider.

I consider two provider payment mechanisms. Under *Prospective Payment System (PPS)*, the payer pays a fixed price  $\Psi^i$  to provider  $i$  for each unit of service provided, irrespective of the service qualities. Under *Pay for Performance (P4P)*, the payer pays a base price  $\Phi^i$ , together with a bonus to provider  $i$  for each unit of service provided. In this section, I assume that both quality dimensions are contractible. When the qualities of provider  $i$ 's services are  $q^i, r^i$ , the bonus per unit of service is  $\alpha^i q^i + \beta^i r^i$ , where  $\alpha^i$  and  $\beta^i$  are the bonus rates chosen by the payer.<sup>12</sup> The contractibility assumption will be relaxed in Subsection 4.3. I allow the payer to make a lump-sum transfer  $\Gamma^i$  to provider  $i$  under both regimes.

The payer assigns equal weight to consumers' payoffs and its own payoff, but a lower weight to the providers' profits.<sup>13</sup> Therefore, its objective is to implement the first best defined in Lemma 1 at the lowest cost. In the extensive form game, the payer sets the copayments and provider payment mechanism in Stage 1.<sup>14</sup> The providers then choose qualities non-cooperatively and simultaneously in Stage 2. The first best is implemented if it is the sub-game perfect Nash equilibrium outcome.<sup>15</sup>

---

<sup>12</sup>For simplicity, I restrict attention to bonuses that are linear in qualities throughout this paper. This approach is in line with a number of theoretical studies on pay for performance in healthcare. See, for example, Eggleston (2005), Kaarboe and Siciliani (2011), and Sherry (2016). In practice, both the U.K. National Health Service and the U.S. Medicare reward providers according to the number of points the providers have earned in various quality domains.

<sup>13</sup>The assumption that the payer assigns a lower weight to the providers' profits is standard. See, for example, Baron and Myerson (1982). The assumption that the payer assigns an equal weight to consumers' payoffs and its own payoff is made to ease notation. Introducing social cost of public fund does not increase the number of margins in the first best, and hence, will not alter implementation results in the following analysis.

<sup>14</sup>This determines a proper subgame in the continuation game.

<sup>15</sup>In Appendix B, I provide sufficient conditions to guarantee that a subgame in Stage 2 has a unique pure-strategy Nash equilibrium.

The next subsection derives the optimal copayments. Subsections 3.2 and 3.3 characterize the complementary optimal provider payment mechanisms.

### 3.1 Optimal Copayments

To implement the first best, the payer has to induce providers to choose the first-best qualities, and induce consumers to follow the first-best allocation. I study the implementation problem in two steps. In the first step, I let each provider  $i$  choose the first-best qualities and derive the optimal copayments that achieve the first-best consumer allocation.

**Lemma 3** *Under managed competition, the first best is implementable only if the copayments satisfy*

$$\Lambda_*^i - \Lambda_*^j = C_D^i(D_*^i, q_*^i, r_*^i) - C_D^j(D_*^j, q_*^j, r_*^j). \quad (9)$$

Recall that under unmanaged competition, the first best can be achieved only if the coinsurance rate is one, so that consumers incur the full service costs. Lemma 3, on the other hand, shows that it is possible to implement the first best without making consumers bear the full costs. In particular, the lemma shows that efficient consumer allocation can be achieved as long as the difference in copayments,  $\Lambda_*^i - \Lambda_*^j$ , are set equal to the difference in the first-best marginal service costs, which the social planner uses to determine the first-best consumer allocation in Lemma 1.<sup>16</sup> This is because consumer choices in a covered market are driven by the price differential rather than the absolute prices. This gives the payer one degree of freedom in setting copayments under managed competition.

The degree of freedom for efficiency concern, together with risk-spreading concern (that is not explicitly modeled), support the use of reference pricing in health insurance design. Under referencing pricing policies, consumers pay a lower copayment for using lower-cost drugs or providers. Condition (9) shows that the efficient copayment for the low-marginal-cost provider can be as low as zero. Moreover, the efficient incremental copayment for the high-marginal-cost provider should be set equal to the incremental marginal service cost, rather than the incremental average cost. The same principle shows that in order for Preferred Provider Organization (PPO) to achieve allocative efficiency, the difference between the in-network and out-of-network cost shares should reflect the difference in marginal costs. Finally, the consumer quality benefits are excluded in the efficient copayments in (9) because consumers have taken benefits into account when they choose providers for services.

---

<sup>16</sup>In this paper, I will focus on optimal compayments for simplicity. However, any optimal consumer cost share policy that implements the first best has to satisfy (9) in equilibrium. For the optimal design of affine linear consumer cost shares that consist of copayment and coinsurance rates under imperfect competition, see Bardey et al. (2016).

### 3.2 Optimal Prospective Payment System

In addition to copayments, the payer also sets prospective prices  $\Psi_*^i$  in Stage 1 under PPS. The payer implements the first best if given the copayments and prospective prices, the first-best qualities are the equilibrium choices of each provider  $i$ , whose maximization program in Stage 2 is

$$\max_{q^i, r^i} \left[ \frac{1}{2} + \frac{v_q(q^i - q^j) + v_r(r^i - r^j) - \{\Lambda_*^i - \Lambda_*^j\}}{2\tau} \right] \Psi_*^i - C^i(D^i, q^i, r^i) + \Gamma_*^i, \quad (10)$$

where  $\Gamma_*^i$  is a lump-sum transfer.

**Proposition 1** *Suppose provider  $i$ 's maximization program is given by (10) and  $\Lambda_*^i$  satisfies Lemma 3. Under Prospective Payment System, the payer implements the first best by choosing*

$$\Psi_*^i = 2\tau D_*^i + C_D^i(D_*^i, q_*^i, r_*^i) = \tau + [v_q(q_*^i - q_*^j) + v_r(r_*^i - r_*^j)] + C_D^j(D_*^j, q_*^j, r_*^j). \quad (11)$$

Moreover,  $\Gamma_*^i$  can be chosen such that provider  $i$ 's equilibrium profit is zero.

The Proposition says that by using two prospective prices and one cost share differential, the payer can implement the first best allocation defined in Lemma 1, which comprises five choice variables.<sup>17</sup> And that two more lump-sum payments are needed for profit extractions. The last subsection has characterized the optimal copayments. Now consider provider  $i$ 's quality choice problem in (10). For each quality dimension  $k$ ,  $k = q, r$ , the marginal social benefit and marginal gross profit of provider  $i$ 's service are, respectively,

$$\underbrace{D^i \times v_k}_{\text{market share} \times \text{marginal benefit}} \quad \text{and} \quad \underbrace{v_k/2\tau \times [\Psi^i - C_D^i(D^i, q^i, r^i)]}_{\text{marginal market share} \times \text{markup}}.$$

Observe that the marginal benefit  $v_k$  in both terms can be cancelled out to make the residual difference independent of quality  $k$ . The independence makes it feasible to equalize the two terms by a single prospective price. Hence, the payer can use (11) to induce each profit-maximizing provider to choose two qualities optimally. In fact, the value of  $\Psi^i$  is the same as the price that each provider  $i$  chooses under unmanaged competition when the coinsurance rate in Subsection 2.3 is equate to one. However, the prospective price is now separated from the copayments, and hence, they have no direct impacts on consumer choices.

The expression for  $\Psi_*^i$  on the right-hand side of (11) does not depend on each provider  $i$ 's own marginal service cost. This is because any increase in  $C_D^i(D_*^i, q_*^i, r_*^i)$  is offset by a proportional

---

<sup>17</sup>Introducing cost reduction choices *à la* Ma (1994) and Hart et al. (1997) into the model does not affect the first-best implementation result. See Appendix B for the details.

decrease in the optimal markup  $2\tau D_*^i$  (see the definition of  $D_*^i$  in (2)).<sup>18</sup> Hence,  $\Psi_*^i$  is based on the marginal service cost of provider  $j$  only, and may first appear to resemble the optimal price in Shleifer’s (1985) yardstick competition model, the model that forms a theoretical basis of prospective payment systems in use. However, there are important differences between Shleifer’s yardstick price and  $\Psi_*^i$ . First, Shleifer’s providers are identical local monopolists supplying services with exogenous, one-dimensional, and identical quality, whereas the providers here compete by supplying endogenous and multiple qualities. The endogeneity accounts for the squared-bracket term in (11), which rewards a provider if consumers value its services more, and *vice versa*. Second, the  $\tau$  in (11) is a markup that a provider gets from consumers’ horizontal preferences, *à la* Hotelling. Since  $\tau$  can be interpreted as a measure of market competitiveness, (11) states that the optimal prospective prices are decreasing in the degree of provider competition.<sup>19</sup>

I will later show that PPS becomes less capable in more complicated environments. However, Proposition 1 still provides some useful guidance on optimal payment design. In practice, the U.S. Medicare’s inpatient prospective prices are based on the average service costs, and are adjusted for factors such as local wage index and medical education (Krinsky et al., 2016). Proposition 1 identifies three potential sources of inefficiency in the current system. First, the system may overpay services that have marginal costs lower than average costs, and *vice versa* (see Table 1 in the next subsection for examples). Second, the system does not take quality differences into account, so it may overpay low-quality hospitals and underpay high-quality ones. Third, the system does not adjust for service market competitiveness, so it may overpay hospitals in more competitive hospital markets, and *vice versa*.

### 3.3 Optimal Pay for Performance

Under P4P, the payer chooses base price  $\Phi_*^i$ , and bonus rates  $\alpha_*^i, \beta_*^i$  in addition to copayment for each provider  $i$ . The maximization program of provider  $i$  is

$$\max_{q^i, r^i} \left[ \frac{1}{2} + \frac{v_q(q^i - q^j) + v_r(r^i - r^j) - \{\Lambda_*^i - \Lambda_*^j\}}{2\tau} \right] [\Phi_*^i + \alpha_*^i q^i + \beta_*^i r^i] - C^i(D^i, q^i, r^i). \quad (12)$$

The first best is implemented if the first-best qualities are the equilibrium choice of provider  $i$ .

---

<sup>18</sup>When the providers have symmetric cost structure so that the subscript can be suppressed, (11) becomes  $\Psi_* = \tau + C_D(1/2, q_*, r_*)$ . The results below can be adjusted in the same way to accommodate symmetric cost structure.

<sup>19</sup>Moreover, Shleifer’s yardstick prices are determined *ex post* by the unit costs chosen by shadow providers; whereas the value of  $\Psi_*^i$  is determined by the regulator *ex ante*, as in most of the managed healthcare markets in practice. This suggests that managing imperfect quality competition is informationally more demanding than regulating local monopolies.

**Proposition 2** Suppose provider  $i$ 's maximization program is given by (12) and copayments satisfy Lemma 3. Under Pay for Performance, the payer implements the first best by choosing  $\alpha_*^i$  and  $\beta_*^i$  such that

$$\alpha_*^i q_*^i = v_q q_*^i - \varepsilon_{q_*}^i \left\{ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - C_D^i(D_*^i, q_*^i, r_*^i) \right\} \quad (13)$$

$$\beta_*^i r_*^i = v_r r_*^i - \varepsilon_{r_*}^i \left\{ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - C_D^i(D_*^i, q_*^i, r_*^i) \right\} \quad (14)$$

where  $\varepsilon_{k_*}^i = \frac{v_k k_*^i}{2\tau D_*^i}$  is provider  $i$ 's quality elasticity of demand at  $k_*^i$ ,  $k = q, r$ , and  $\Phi_*^i$  is chosen such that provider  $i$ 's equilibrium profit is zero.

It is of course not surprising that the first best can be implemented by contracting on qualities. However, Proposition 2 will serve as a benchmark in the analysis that follows. The construction of optimal bonuses in the proposition also deserves attention. Consider (13). The bonus rate  $\alpha_*^i$  affects provider  $i$ 's incentive to increase  $q^i$  through two channels. First, the bonus rate is the marginal revenue that provider  $i$  obtains from raising the quality. First-best implementation requires marginal private benefit to be aligned with marginal social benefit. This accounts for the term  $v_q q_*^i$  in (13).

However, provider  $i$ 's marginal private benefit of choosing  $q_*^i$  is different from  $v_q q_*^i$ . This is due to the second channel, demand response, which accounts for the term  $\varepsilon_{q_*}^i \{\cdot\}$  in (13). When the bonus induces provider  $i$  to raise quality, the provider attracts more consumers from its competitor. The demand responsiveness is captured by  $\varepsilon_{q_*}^i$ , the quality elasticity of demand. For each additional consumer, provider  $i$ 's incremental revenue is  $\Phi_*^i + \alpha_*^i q^i + \beta_*^i r^i$ . In order to extract all profits from provider  $i$ , the payer sets the base price  $\Phi_*^i$  such that the incremental revenue equals the average service cost  $C^i(D_*^i, q_*^i, r_*^i)/D_*^i$ . However, provider  $i$ 's marginal service cost is  $C_D^i(D_*^i, q_*^i, r_*^i)$ . If the marginal cost is below the average cost, the provider enjoys a positive markup, and hence, it has excessive incentives to attract consumers by raising quality  $q^i$ . Therefore, the term  $\varepsilon_{q_*}^i \{\cdot\}$  is subtracted from  $v_q q_*^i$  in (13) to implement the first best. The reverse is true when the marginal cost is above the average cost, so that the markup is negative. The value of  $\beta_*^i r_*^i$  in (14) can be interpreted in the same way.

Technology	Example	Optimal bonus rate
Marginal cost < Average cost	Imaging service	Lower than marginal social benefit and decreasing in demand elasticity
Marginal cost = Average cost	Drug therapy	The same as marginal social benefit and independent of demand elasticity
Marginal cost > Average cost	Surgery	Higher than marginal social benefit and increasing in demand elasticity

Table 1: Optimal bonus rate

Under PPS, the optimal prospective prices are always increasing in market competitiveness (or decreasing in  $\tau$ ). However, conditions (13) and (14) show that the optimal bonus rates can be increasing or decreasing in demand response, which is proportional to market competitiveness. Table 1 illustrates how the relationship between bonus rate and demand response is determined by a service’s cost structure. For example, the optimal bonus rates for high-fixed-cost imaging service can be very close to zero when consumers are responsive to service qualities.<sup>20</sup> The intricate relationships in the table also provide a potential explanation for the mixed provider responses to uniform bonus rates across services and markets in practice. Finally, the value of the optimal base price  $\Phi_*^i$  in the proposition is given by (42) in Appendix A. Because  $\Phi_*^i$  is sufficient to implement the first best at zero profit, lump-sum transfer is not needed.

## 4 Consumer misperceptions, not-for-profit providers, non-contractible quality, and alternative market structures

In this section, I relax the assumptions that consumers are perfectly rational and providers are purely profit-maximizing. I first consider implementation problems in which the payer has the same instruments as in the last subsection. I will then relax the assumption that qualities are perfectly contractible in Subsection 4.3. Finally, Subsection 4.4 will consider two alternative market structures in which providers do not compete with each other.

### 4.1 Consumer misperceptions and prices

The analysis so far has assumed that  $v_q q^i + v_r r^i$  is both the true and perceived service benefits. However, the medical literature has shown that patients can have wrong perceptions about providers and services.<sup>21</sup> I now incorporate consumer misperceptions into the model. I continue to let  $v_q q^i + v_r r^i$  be a consumer’s true benefit from provider  $i$ ’s service. However, the same consumer perceives that the benefit from provider  $i$ ’s service is  $\tilde{v}_q^i q^i + \tilde{v}_r^i r^i$ , where every  $\tilde{v}_k^i$ ,  $k = q, r$ , is positive

---

<sup>20</sup>Moreover, (13) and (14) show that the payer can ignore the cross effects between qualities of the same provider or competing provider. And this is true even if the quality benefits are not additive separable in the consumers’ benefit function.

<sup>21</sup>At the provider level, Doyle et al. (2013) show that the association between patient experience and technical quality of care is mixed in the medical literature. However, there are more consistent associations between patient experience and some other measures of clinical quality. At the service level, Schwartz et al. (2014) find that at least 25% of the U.S. Medicare beneficiaries have received low-value services that provide little to no clinical benefits in 2009. This observation cannot be explained by *ex post* moral hazard if consumers are perfectly rational.

and provider specific.<sup>22</sup> The consumer overestimates the benefit of provider  $i$ 's quality  $k^i$  if  $\tilde{v}_k^i > v_k$ , and underestimates the benefit if  $v_k > \tilde{v}_k^i > 0$ .

The payer seeks to implement the first best characterized in Lemma 1, which is based on consumer true service benefits. On the other hand, consumer choices are based on perceived service benefits.<sup>23</sup> The payer chooses copayments to influence consumer choices as in Subsection 3.1. The following lemma shows how the optimal copayments in Lemma 3 have to be adjusted to accommodate consumer misperceptions.<sup>24</sup>

**Lemma 4** *Suppose consumers' true and perceived benefits from provider  $i$ 's service are  $v_q q^i + v_r r^i$  and  $\tilde{v}_q^i q^i + \tilde{v}_r^i r^i$ , respectively. Under managed provider competition, the first best is implementable only if the payer sets*

$$\tilde{\Lambda}_*^i - \tilde{\Lambda}_*^j = C_D^i(D_*^i, q_*^i, r_*^i) - C_D^j(D_*^j, q_*^j, r_*^j) + \left[ \{(\tilde{v}_q^i - v_q)q_*^i + (\tilde{v}_r^i - v_r)r_*^i\} - \{(\tilde{v}_q^j - v_q)q_*^j + (\tilde{v}_r^j - v_r)r_*^j\} \right]. \quad (15)$$

The squared-bracket term in (15) complements the cost-based reference pricing rule by requiring consumers to internalize the difference between perceived and true quality benefits. This forms the basis of "value-based" insurance design when consumers misperceive the benefits from using different providers or services. For example, suppose consumers overestimate the benefit of quality  $q$  offered by provider  $i$ , so that  $\tilde{v}_q^i > v_q$ . Condition (15) says that the payer should raise the cost-based copayment differential in Lemma 3 by  $(\tilde{v}_q^i - v_q)q_*^i$  to discourage consumers from overutilizing provider  $i$ ' service. Now suppose consumers overestimate the benefits of quality  $q$  offered by both providers, so that  $\tilde{v}_q^i = \tilde{v}_q^j = \tilde{v}_q > v_q$ . The squared-bracket term in (15) becomes  $(\tilde{v}_q - v_q)(q_*^i - q_*^j)$ , which is positive when  $q_*^i > q_*^j$ . In general, the copayment differential in (15) can accommodate multiple sources of consumer misperceptions so long as the market is covered.<sup>25</sup> Thus, the payer can retain the degree of freedom in setting copayments, as in Lemma 3.

---

<sup>22</sup> Again, all the results continue to hold if the perceived benefit is  $\tilde{V}^i(q^i, r^i)$ , where  $\tilde{V}^i$  is provider specific, twice differentiable, strictly increasing, and concave.

<sup>23</sup> In the terminology of Kahneman et al. (1997),  $v_q q^i + v_r r^i$  and  $\tilde{v}_q^i q^i + \tilde{v}_r^i r^i$  are the consumers' "experience utilities" and "decision utilities", respectively. Welfare is based on experience utilities but consumer choices are based on decision utilities.

<sup>24</sup> The difference between  $v_q q^i + v_r r^i$  and  $\tilde{v}_q^i q^i + \tilde{v}_r^i r^i$  represents the (residual) misperceptions that cannot be eliminated by information "nudge" (see, for example, Thaler and Sunstein, 2008). The mechanism design approach I use here is consistent with the idea of optimal paternalism in O'Donoghue and Rabin (2003). See Farhi and Gabaix (2015) for an application of this approach to the design of optimal taxation.

<sup>25</sup> The first best would not be implementable by a single price differential if the misperceptions were heterogeneous among consumers. Armstrong (2015) shows that when there are both rational and misperceived consumers in the same market, the misperceived consumers may cross-subsidize the rational consumers in equilibrium.



## 4.2 Providers' non-pecuniary benefits and profit-distribution constraints

Other than perfectly rational consumers, the analysis in Section 3 has also assumed that providers are purely profit-maximizing. However, many healthcare professionals enjoy non-pecuniary benefits from providing quality services. Moreover, many institutional providers are not-for-profit organizations. Equipped with Lemma 4, I now study the payer's payment design problem when consumers have quality misperceptions, and providers have intrinsic quality incentives.

From now on, I assume that provider  $i$ , either as an individual or an organization's executive, receives non-pecuniary benefit from providing services. These non-pecuniary benefits can be broadly defined to include genuine altruism and warm-glow benefit such as social status and recognition. When provider  $i$  serves a consumer at quality levels  $(q^i, r^i)$ , the non-pecuniary benefit is  $\eta_q^i q^i + \eta_r^i r^i$ , where  $\eta_q^i$  and  $\eta_r^i$  are provider specific and strictly positive.<sup>26</sup> I assume that the non-pecuniary benefits cannot be monetized or extracted. Moreover, the non-pecuniary benefits do not enter the social welfare function, so Lemma 1 continues to characterize the first best.<sup>27</sup> I also allow provider  $i$  to be a not-for-profit organization. I follow Glaeser and Shleifer (2001) to model non-profit status as a constraint on profit distribution. The constraint forces a non-profit organization's executives to spend profits on perquisites, which are less valuable than cash. Therefore, when provider  $i$  is a not-for-profit organization and its profit is  $\Pi^i$ , the executives' total payoff from perquisites is only  $\theta^i \Pi^i$ , where  $1 > \theta^i > 0$ . However, I will relax the range of  $\theta^i$  to  $1 \geq \theta^i > 0$  in the following analysis in order to consider both for-profit and not-for-profit providers who enjoy non-pecuniary benefits.<sup>28</sup>

Under PPS, the non-pecuniary benefit and non-distribution constraint, together with consumer misperceptions, modify provider  $i$ 's maximization program (10) in Subsection 3.2 to

$$\max_{q^i, r^i} \left[ \frac{1}{2} + \frac{\{\tilde{v}_q^i q^i + \tilde{v}_r^i r^i\} + \{\tilde{v}_q^j q^j + \tilde{v}_r^j r^j\} - \{\tilde{\Lambda}_*^i - \tilde{\Lambda}_*^j\}}{2\tau} \right] \left[ \{\eta_q^i q^i + \eta_r^i r^i\} + \theta^i \tilde{\Psi}_*^i \right] - \theta^i C^i(\tilde{D}^i, q^i, r^i) + \theta^i \tilde{\Gamma}_*^i, \quad (16)$$

where the first squared-bracket term is the definition of provider  $i$ 's market demand  $\tilde{D}^i$  when consumers are misperceived. These modifications significantly limit the capability of PPS, as the following proposition shows.

**Proposition 3** *Suppose provider  $i$ 's maximization program is given by (16) and  $\tilde{\Lambda}_*^i$  satisfies Lemma*

<sup>26</sup>The non-pecuniary benefit can be generalized to  $\eta^i(q^i, r^i)$ , where  $\eta^i$  is twice differentiable, strictly increasing, and concave, without changing the results.

<sup>27</sup>I exclude the firms' non-pecuniary benefits from social welfare to avoid the problem of double counting, see, for example, Hammond (1987) and Milgrom (1993). Moreover, I assume that perquisites have no social benefit.

<sup>28</sup>In Glaeser and Shleifer (2001), the entrepreneur suffers a non-pecuniary cost of quality shirking, no matter that the firm is for-profit or not-for-profit.

4. Under Prospective Payment System, the payer can implement the first best only if

$$\frac{v_q - \eta_q^i/\theta^i}{v_r - \eta_r^i/\theta^i} = \frac{\tilde{v}_q^i}{\tilde{v}_r^i}. \quad (17)$$

When (17) holds, the first best can be implemented by

$$\tilde{\Psi}_*^i = \left[ 2\tau D_*^i \left( \frac{v_r - \eta_r^i/\theta^i}{\tilde{v}_r^i} \right) - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \right] + C_D^i(D_*^i, q_*^i, r_*^i), \quad (18)$$

and  $\tilde{\Gamma}_*^i$  can be chosen such that provider  $i$ 's equilibrium profit is zero.

While PPS implements the first best unconditionally in Subsection 3.2, it implements the first best here only if (17), a knife-edge condition, holds. On the left-hand side of (17) is the consumer marginal rate of substitution (MRS) required for provider  $i$  to choose the first-best quality mix under PPS. When provider  $i$  is profit-maximizing, the ratio is simply  $v_q/v_r$ . However, both the non-pecuniary benefit and non-distribution constraint incentivize the provider to raise qualities. Therefore, the required demand response of quality  $k^i$  is reduced by the level of  $\eta_k^i/\theta^i$ , which adjusts the required MRS to the ratio in (17). On the right-hand side of (17) is the actual MRS. For perfectly rational consumers, the ratio is again  $v_q/v_r$ , but consumer misperceptions change it to  $\tilde{v}_q^i/\tilde{v}_r^i$ . Because of the changes on both sides of (17), the first best is generically not implementable when providers have intrinsic quality incentives and/or consumers are misperceived.

For the sake of completeness and future reference, consider the optimal prospective price  $\tilde{\Psi}_*^i$  when (17) holds. This is (18) in the proposition. Recall that in Subsection 3.2, the markup in the prospective price is  $2\tau D_*^i$ . That markup is adjusted to the squared-bracket term in (18), which is decreasing in provider  $i$ 's non-pecuniary benefit, and cost of non-distribution constraint (due to a smaller  $\theta^i$ ).<sup>29</sup> The markup is also decreasing in consumers' misperceived quality benefit.

Now turn to P4P. Provider  $i$ 's preference and constraint, together with consumer misperceptions turn maximization program (12) in Subsection 3.3 to

$$\max_{q^i, r^i} \left[ \frac{1}{2} + \frac{\{\tilde{v}_q^i q^i + \tilde{v}_r^i r^i\} - \{\tilde{v}_q^j q_*^j + \tilde{v}_r^j r_*^j\} - \{\tilde{\Lambda}_*^i - \tilde{\Lambda}_*^j\}}{2\tau} \right] \left[ \{\eta_q^i q^i + \eta_r^i r^i\} + \theta^i \left\{ \tilde{\Phi}_*^i + \tilde{\alpha}_*^i q^i + \tilde{\beta}_*^i r^i \right\} \right] - \theta^i C^i(\tilde{D}^i, q^i, r^i). \quad (19)$$

However, the optimal bonuses can be adjusted to accommodate these new elements.

**Proposition 4** Suppose provider  $i$ 's maximization program is given by (19) and  $\tilde{\Lambda}_*^i$  satisfies Lemma 4. Under Pay for Performance, the payer implements the first best by choosing  $\tilde{\alpha}_*^i$  and  $\tilde{\beta}_*^i$  such that

$$\tilde{\alpha}_*^i q_*^i = \left[ v_q - \frac{\eta_q^i}{\theta^i} \right] q_*^i - \tilde{\varepsilon}_{q_*}^i \left\{ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - \left[ C_D^i(D_*^i, q_*^i, r_*^i) - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \right] \right\} \quad (20)$$

$$\tilde{\beta}_*^i r_*^i = \left[ v_r - \frac{\eta_r^i}{\theta^i} \right] r_*^i - \tilde{\varepsilon}_{r_*}^i \left\{ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - \left[ C_D^i(D_*^i, q_*^i, r_*^i) - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \right] \right\} \quad (21)$$

<sup>29</sup>Using the definition of  $D_*^i$ , it can be shown that  $\tilde{\Psi}_*^i$  is generically not independent of  $C_D^i(D_*^i, q_*^i, r_*^i)$ .

where  $\tilde{\varepsilon}_{k_*}^i = \frac{\tilde{v}_k^i k_*^i}{2\tau D_*^i}$  is provider  $i$ 's quality elasticity of demand at  $k_*^i$ ,  $k = q, r$ . Moreover,  $\tilde{\Phi}_*^i$  is chosen such that provider  $i$ 's equilibrium profit is zero.

Consider the comparison between (20) and (13). The changes in consumer response and provider objective introduce three adjustments to the optimal bonuses. First, because consumers are influenced by their misperceptions, the demand elasticity in (20) is updated to  $\tilde{\varepsilon}_{q_*}^i$ . Second, both Provider  $i$ 's (marginal) non-pecuniary benefit and non-distribution constraint act as direct substitutes of explicit quality incentive. This accounts for the first squared-bracket term in (20). Finally, the (total) non-pecuniary benefit and non-distribution constraint also induce provider  $i$  to expand its market share, as if it has a lower marginal service cost. This accounts for the second squared-bracket term in (20). The optimal  $\tilde{\beta}_*^i$  in (21) is adjusted in the same way.<sup>30</sup>

Both the optimal prospective price in Proposition 3 and the optimal bonuses in Proposition 4 show that consumer misperception is a key determinant of optimal provider payments (whereas the optimal copayments do not depend on providers' objectives). The payment formulae also confirm that providers' non-pecuniary benefits can reduce the magnitude of monetary incentives (Besley and Ghatak, 2005), and that non-distribution constraint can incentivize quality provision (Glaeser and Shleifer, 2001). Indeed, Glaeser and Shleifer (2001), Besley and Ghatak (2005), and a large follow-up literature have used single-quality models to argue that non-pecuniary benefit and non-distribution constraint are substitutes of explicit incentives.

The primary implication of Propositions 3 and 4, however, is that when service quality is multi-dimensional, explicit incentives are complements of non-pecuniary benefit and non-distribution constraint. In particular, (17) shows that under PPS both non-pecuniary benefit and non-distribution constraint distort a provider's marginal private quality benefit away from the marginal social quality benefit. This provider paternalism can be corrected by targeted quality bonuses but not generic prices. This suggests that in health service markets, where providers generally possess intrinsic quality incentives, P4P has a less significant role in incentivizing qualities, but a more significant role in directing resource allocation.

### 4.3 Non-contractible quality

Until now the two qualities are contractible, and P4P can always perform weakly better than PPS. I now relax the perfect contractibility assumption. I follow Benabou and Tirole (2016) to assume that one of the two qualities, quality  $r$ , is not contractible, because the quality is not observable/measurable by the payer. Other settings are the same as the previous two subsections.

---

<sup>30</sup>The values of  $\tilde{\alpha}_*^i$  and  $\tilde{\beta}_*^i$  in the proposition can be negative. Suppose that the regulator is confined to set positive bonus rates, and that  $\tilde{\beta}_*^i < 0$ . Proposition 5 in the next subsection shows that the payer can still implement the first best if a sufficient condition holds.

In particular, consumers continue to observe the two qualities of each provider, but their perceptions about service benefits can be wrong.<sup>31</sup>

Consider PPS. Here, provider  $i$  receives a fixed price per unit of service, irrespective of the levels of  $q^i$  and  $r^i$ . Therefore, the non-contractability of  $r^i$  has no impact on the implementability of the first best and the optimal prospective prices. I state this obvious result as a corollary.

**Corollary 1** *Proposition 3 remains valid when quality  $r$  is non-contractible.*

How about P4P? When  $r^i$  is not contractible, the payer can only use base price and bonus rate  $\hat{\alpha}_*^i$  to incentivize provider  $i$ . If the payer can also make a lump-sum transfer to provider  $i$ , the provider's maximization program becomes

$$\max_{q^i, r^i} \left[ \frac{1}{2} + \frac{\{\tilde{v}_q^i q^i + \tilde{v}_r^i r^i\} - \{\tilde{v}_q^j q^j + \tilde{v}_r^j r^j\} + \{\tilde{\Lambda}_*^i - \tilde{\Lambda}_*^j\}}{2\tau} \right] \left[ \{\eta_q^i q^i + \eta_r^i r^i\} + \theta^i \left\{ \hat{\Phi}_*^i + \hat{\alpha}_*^i q^i \right\} \right] - \theta^i C^i(\tilde{D}^i, q^i, r^i) + \theta^i \hat{\Gamma}_*^i. \quad (22)$$

It turns out that these instruments can also implement the first best at zero profit if a sufficient condition holds.

**Proposition 5** *Suppose provider  $i$ 's maximization program is given by (22) and  $\tilde{\Lambda}_*^i$  satisfies Lemma 4. Furthermore, suppose that the left-hand side of (17) is larger than its right-hand side. Under Pay for Performance, the payer implements the first best by choosing  $\hat{\alpha}_*^i$  and  $\hat{\Phi}_*^i$  such that  $\hat{\Phi}_*^i = \tilde{\Psi}_*^i - \hat{\alpha}_*^i q_*^i$ , where  $\tilde{\Psi}_*^i$  is given by (18) in Proposition 3 and where*

$$\hat{\alpha}_*^i = \left[ v_q - \frac{\eta_q^i}{\theta^i} \right] - \frac{\tilde{v}_q^i}{\tilde{v}_r^i} \left[ v_r - \frac{\eta_r^i}{\theta^i} \right] \quad (23)$$

*is strictly positive. Moreover,  $\hat{\Gamma}_*^i$  can be chosen such that provider  $i$ 's equilibrium profit is zero.*

In Proposition 4, where both qualities are contractible, each of the two bonus rates is set so that provider  $i$ 's marginal private benefit from each quality is the same as the marginal social benefit at the first best. But  $\hat{\alpha}_*^i$  cannot be set in that way when  $r$  is non-contractible. In fact, this multi-tasking problem has been studied extensively in the bilateral setting. In this literature, the optimality of rewarding the contractible qualities depends on whether the contractible and non-contractible qualities are substitutes and complements in the principal's benefit function, and the agent's production function (Holmstrom and Milgrom, 1991; Kaarboe and Siciliani, 2011; Sherry, 2016). However, Proposition 5 shows that these relationships are inconsequential here.<sup>32</sup> This is

<sup>31</sup>If quality  $r$  is observable by the payer but not the consumers, the payer can inform the consumers by reporting the quality publicly. Consumer imperfect inferences from the report can be captured by  $\tilde{v}_r^i$ .

<sup>32</sup>This is true even if consumers true and perceived benefits are  $V(q^i, r^i)$  and  $\tilde{V}^i(q^i, r^i)$ . This is because Proposition 5 is based on first derivatives (marginal benefits) but not cross-partial derivatives. Whether the two qualities are substitutes or complements in consumption are inconsequential.

because under managed competition, provider  $i$  is also responding to consumers' quality preferences, which are absent in the bilateral framework.

In order to understand how the value of  $\hat{\alpha}_*^i$  in Proposition 5 is determined, consider Proposition 3 again. There, PPS fails to implement the first best whenever (17) does not hold. Now let the left-hand side of (17) to be larger than the right-hand side, so that consumers' preference for quality  $q$  is weaker than what is required to implement the first best under PPS. Fix provider  $i$ 's revenue from a consumer at  $\tilde{\Psi}_*^i$  in Proposition 3, and introduce the bonus rate  $\hat{\alpha}_*^i$  in Proposition 5 to the provider's maximization problem. The bonus incentivizes the provider to raise quality  $q^i$ , and reduces the required demand response of  $q^i$  by the level of  $\hat{\alpha}_*^i$ . Hence, the required MRS on the left-hand side of (17) becomes

$$\frac{v_q - \eta_q^i/\theta^i - \hat{\alpha}_*^i}{v_r - \eta_r^i/\theta^i} = \frac{v_q - \eta_q^i/\theta^i - [(v_q - \eta_q^i/\theta^i) - (\tilde{v}_q^i/\tilde{v}_r^i)(v_r - \eta_r^i/\theta^i)]}{v_r - \eta_r^i/\theta^i} = \frac{\tilde{v}_q^i}{\tilde{v}_r^i},$$

which is identical to the actual MRS on the right-hand side of (17). Therefore, the first best becomes implementable. The implementation still requires that the provider's revenue from a consumer be  $\tilde{\Psi}_*^i$ , but this can be accomplished by the optimal base price  $\hat{\Phi}_*^i = \tilde{\Psi}_*^i - \hat{\alpha}_*^i q_*^i$ .

What if the left-hand side of (17) is smaller than the right-hand side? Here, consumers' preference for quality  $q$  is stronger than what is required to implement the first best under PPS. In principle, the payer can use a negative  $\hat{\alpha}_*^i$  that satisfies (23) to implement the first best. However, quality penalty is practically infeasible to implement, whereas any positive bonus will distort qualities further away from the first best. Thus, PPS does perform better than P4P in this case. These results suggest that under partial contractibility, the optimal choice between PPS and P4P is determined by consumers' preferences over contractible and non-contractible qualities.<sup>33</sup>

#### 4.4 Alternative market structures

In order to isolate the effects of imperfect competition on copayment and provider payment design, this subsection considers two closely related market structures where competition is suppressed. In the first scenario, the market share of provider  $i$  is exogenously fixed at  $\bar{D}^i$  due to capacity constraints or other restrictions.<sup>34</sup> In the first best, qualities  $q_*^i$  and  $r_*^i$  are still defined by (3) and (4) after  $D_*^i$  in those conditions are replaced by  $\bar{D}^i$ . However, without genuine consumer choices, copayments become lump-sum transfers that do not affect qualities. On the supply side, the variable  $\tilde{D}^i$  in the maximization programs in this section has to be replaced by the constant  $\bar{D}^i$ . Under

---

<sup>33</sup>In the setting of hip replacement surgery in the English National Health Service, Gutaker et al. (2016) study the strength of patients' preferences over different quality measures.

<sup>34</sup>Chalkley and Malcomson (1998a) is the first paper to study optimal contracting without demand response, where quality is unidimensional, and provider engages in cost-reduction activities.

PPS, the first best is implementable only if  $v_q/v_r = \eta_q^i/\eta_r^i$ , so that the payer and provider  $i$  share the same preference over qualities. Under P4P with full contractibility, (20) and (21) continue to characterize the bonus rates that implement the first best unconditionally, after all  $\tilde{\varepsilon}_{k^*}^i$  in the two conditions are set to zero, so that the effects of competition and technology on bonus rates are muted. When quality  $r$  is not contractible, P4P can implement the first best only if  $v_q/v_r > \eta_q^i/\eta_r^i$ , so that the payer prefers quality  $q$  more than provider  $i$ . Consumers' preference is inconsequential on both sides of the market.

In the second scenario, only provider  $A$  is active and the Hotelling line is not covered. In this monopoly model, consumers choose between provider  $A$ 's service and no service, which yields zero payoff. In Appendix B, I characterize the first best in which the marginal consumer receives zero (net) social welfare from provider  $A$ 's service. I then consider the first-best implementation problems. On the supply side, the formulae in Propositions 3 to 5 continue to pin down the necessary conditions and optimal payments for implementation, once the demand responses and the corresponding elasticities are adjusted to account for the change in market structure.<sup>35</sup> On the demand side, however, efficient allocation requires consumers to incur the full marginal cost. This shows that the referencing pricing rules in Lemmas 3 and 4 are optimal only when genuine provider choices are available.

## 5 Equity concern

Now return to the managed competition paradigm. The previous sections have studied various first-best implementation problems. But in practice, first-best implementation can be infeasible or undesirable because the payer has concerns other than economic efficiency. This section focuses on equity concern. In particular, I let the payer's new objective be implementing uniform service qualities efficiently. I also let the payer set uniform copayments for the two providers' services. However, I will show that the basic principles for optimal provider payment design apply well beyond the first best.

### 5.1 Second best

In the second best, consumers are perfectly rational and providers are purely profit maximizing as in Section 3. Because there is no quality variation between providers, the market share of each provider is  $1/2$ , and the average transportation is  $\tau/4$ . The social planner, who can dictate qualities,

---

<sup>35</sup>Chalkley and Malcomson (1998b) consider a similar model. They show that when the payer and consumers have the same preference over qualities, the first best can be implemented by a lump-sum transfer and two provider prices, one for the consumers who are served, and one for the consumers who are not served.

chooses uniform  $q$  and  $r$  to maximize social welfare

$$v_q q + v_r r - \tau/4 - C^A(1/2, q, r) - C^B(1/2, q, r).$$

The second-best qualities are characterized by

$$v_q = C_q^A(1/2, q_s, r_s) + C_q^B(1/2, q_s, r_s) \quad (24)$$

$$v_r = C_r^A(1/2, q_s, r_s) + C_r^B(1/2, q_s, r_s). \quad (25)$$

These conditions demonstrate how equity concern reduces production efficiency. When the social planner chooses uniform qualities, it has to balance the marginal social benefit of each quality to the marginal quality costs of both providers. Therefore, providers are generally not producing at their efficient quality levels, where the marginal benefits equal the marginal costs of each provider.

Now turn to the implementation problems. Because equity concern dictates that the payer set uniform copayments (which is optimal when consumers are rational), it is sufficient to revisit the provider payment design problems. The next proposition considers PPS.

**Proposition 6** *Under Prospective Payment System, the second best is implementable only if*

$$\frac{C_q^i(1/2, q_s, r_s)}{C_r^i(1/2, q_s, r_s)} = \frac{v_q}{v_r}. \quad (26)$$

This condition and (17) in Proposition 3 have similar interpretation. In order for PPS to implement either the first best or the second best, the required MRS between qualities must be equal to the actual MRS. For first-best implementation, both ratios are  $v_q/v_r$ . However, the payer's equity concern distorts the second-best qualities from each provider's efficient levels. This distortion introduces marginal quality costs to the required MRS without changing the actual MRS, and renders the second-best qualities generically not implementable.

Proposition 6, together with Propositions 1 and 3, show that PPS can generically implement the payer's ideal allocation only when the payer is maximizing efficiency, providers are maximizing profits, and consumers are perfectly rational. Distortion on any side makes implementation infeasible. I next turn to P4P with both qualities contractible.

**Proposition 7** *Under Pay for Performance, the payer implements the second best by choosing  $\alpha_s^i$  and  $\beta_s^i$  such that*

$$\alpha_s^i q_s = \frac{C_q^i(1/2, q_s, r_s)}{1/2} q_s - \varepsilon_{q_s} \left\{ \frac{C_q^i(1/2, q_s, r_s)}{1/2} - C_D^i(1/2, q_s, r_s) \right\} \quad (27)$$

$$\beta_s^i r_s = \frac{C_r^i(1/2, q_s, r_s)}{1/2} r_s - \varepsilon_{r_s} \left\{ \frac{C_r^i(1/2, q_s, r_s)}{1/2} - C_D^i(1/2, q_s, r_s) \right\} \quad (28)$$

where  $\varepsilon_{k_s} = \frac{v_k k_s}{2\tau 1/2}$  is the common quality elasticity of demand at  $k_s$ ,  $k = q, r$ . Moreover,  $\Phi_s^i$  is chosen such that provider  $i$ 's equilibrium profit is zero.

This result once again illustrates the vigor of P4P. For second-best implementation, the marginal quality benefit,  $v_q$ , in (13) is replaced by the marginal quality cost (per consumer),  $C_q^i(1/2, q_s, r_s)/(1/2)$ , in (27) to accommodate the distortion due to equity concern. This is the only change in the bonus-rate formula. Therefore, the relationship among cost structure, demand elasticity, and bonus rate in Table 1 remains valid. Indeed, the next proposition shows that when quality  $r$  is not contractible, the payer's equity concern also restricts the capability of P4P in a way similar to that due to consumer-side and provider-side distortions.

**Proposition 8** *Suppose that the left-hand side of (26) is larger than its right-hand side. Under Pay for Performance, the payer can implement the second best by choosing  $\hat{\alpha}_s^i$ ,  $\hat{\Phi}_s^i$ , and  $\hat{\Gamma}_s^i$  optimally. In particular,*

$$\hat{\alpha}_s^i = \left[ \frac{C_q^i(1/2, q_s, r_s)}{1/2} \right] - \frac{v_q}{v_r} \left[ \frac{C_r^i(1/2, q_s, r_s)}{1/2} \right]. \quad (29)$$

When the left-hand side of (26) is larger than its right-hand side, the relative marginal cost of  $q_s$  over  $r_s$  is too high for provider  $i$  to choose the second-best quality mix under PPS. Here, the bonus rate in (29) is again chosen to adjust the required MRS, so that the second-best becomes implementable. Nevertheless, P4P cannot outperform PPS when the relative marginal cost of quality  $q_s$  is low in the second best.

## 5.2 Third best

When consumers have provider-specific misperceptions about service benefits but copayments are uniform due to equity concern, neither the first best nor the second best is implementable. In Appendix A, I characterize the third best in which the social planner chooses uniform qualities to maximize social welfare subject to consumer misperceptions. I then consider the third-best implementation problems with the additional assumption that providers are not purely profit-maximizing as in Section 4. I show that Propositions 3 to 5 continue to specify whether the two provider payment systems can implement the third best or not, once adjustments similar to those in the previous subsection are made to account for the discrepancy between marginal quality costs and benefits.

## 6 Concluding remarks

This paper studies a health service market in which explicit incentives, implicit incentives, and market force interact as an incentive system. The main results can be summarized as a practical guide for optimal regulation in the system. First, when consumers are rational and providers are



profit maximizing, PPS and P4P are equivalent. Second, P4P has more flexibility to address misperceived consumers, and motivated but paternalistic providers, when all qualities are contractible. Third, consumers' quality preference determines whether P4P can outperform PPS or not when one of the qualities is non-contractible. Finally, optimal provider payment mechanisms have to be complemented by optimal copayments that are set according to the providers' marginal service costs and consumer misperceptions. These general economic principles can be applied to healthcare markets, as well as markets for nursing homes, childcare, social services, and education.

For tractability, I have assumed that a quality dimension is either perfectly measurable or not measurable at all, and that the payer knows the productivity of each provider. In a bilateral framework, Kuhn and Siciliani (2009) allow the provider to game its quality measure at the margin, and to have private information about its productivity type. They show that when the payer designs a pay-for-performance scheme, it can benefit from pooling different productivity types, and the resulting incentive scheme can be interpreted as a rationale for minimum quality standard. Accommodating gaming and private information in the incentive system studied in this paper is a fruitful direction for future research.

In the domain of healthcare, the model can potentially be extended to incorporate other policy relevant aspects, such as income-based consumer subsidy (e.g., health insurance exchanges in the U.S.), entry and exit of providers (e.g., Medicare Part D markets), and interactions between private providers and a regulator who also provides services (e.g., Medicare Advantage and original Medicare). Designing appropriate regulatory regimes to reconcile social and individual interests in these complicated environments is of practical importance.

## Appendix A

**Proof of Lemma 1:** First, differentiate (1) with respect to  $D^A$  and set the derivative to zero yield (2), which is interior if both providers are active. Second, the convexity of  $C^i$  in  $q^i$  and  $r^i$  guarantees that (1) is concave in  $q^i, r^i$ . Differentiate (1) with respect to  $q^i, r^i$  and set the derivatives to zero yield (3) and (4), respectively.

**Proof of Lemma 2:** Differentiate (5) with respect to  $\Lambda^i, q^i, r^i$ , and set the derivatives to zero yield

$$\frac{D_{\dagger}^i}{\phi} - \frac{1}{2\tau} \left[ \frac{\Lambda_{\dagger}^i}{\phi} - C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i) \right] = 0 \quad (30)$$

$$\frac{v_q}{2\tau} \left[ \frac{\Lambda_{\dagger}^i}{\phi} - C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i) \right] = C_q^i(q_{\dagger}^i, r_{\dagger}^i, s_{\dagger}^i) \quad (31)$$

$$\frac{v_r}{2\tau} \left[ \frac{\Lambda_{\dagger}^i}{\phi} - C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i) \right] = C_r^i(q_{\dagger}^i, r_{\dagger}^i, s_{\dagger}^i). \quad (32)$$

By (30),  $\Lambda_{\dagger}^i = 2\tau D_{\dagger}^i + \phi C_D^i(D_{\dagger}^i, q_{\dagger}^i, r_{\dagger}^i)$ . Substituting this into the squared-bracket term in (5), (31), and (32) give equations (6) to (8). These equations and those in Lemma 1 are identical only if  $D_*^i = 1/2$  and  $\phi = 1$ .

**Proof of Lemma 3:** See proof of Lemma 4.

**Proof of Proposition 1:** See proof of Proposition 3.

**Proof of Proposition 2:** See proof of Proposition 4.

**Proof of Lemma 4:** The first best is implementable only if

$$\begin{aligned} & \frac{1}{2} + \frac{v_q(q_*^i - q_*^j) + v_r(r_*^i - r_*^j) - \{C_D^i(D_*^i, q_*^i, r_*^i) - C_D^j(D_*^j, q_*^j, r_*^j)\}}{2\tau} \\ &= \frac{1}{2} + \frac{\{\tilde{v}_q^i q_*^i + \tilde{v}_r^i r_*^i\} - \{\tilde{v}_q^j q_*^j + \tilde{v}_r^j r_*^j\} - \{\tilde{\Lambda}_*^i - \tilde{\Lambda}_*^j\}}{2\tau}, \end{aligned} \quad (33)$$

which is simplified to (15). Furthermore, (9) in Lemma 3 is obtained by setting  $\tilde{v}_k^i = v_k$ .

**Proof of Proposition 3:** Differentiating (16) with respect to  $q^i, r^i$  and setting the derivatives to zero, I obtain

$$\frac{\tilde{v}_q^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \tilde{\Psi}_*^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \frac{\eta_q^i}{\theta^i} = C_q^i(\tilde{D}^i, q^i, r^i) \quad (34)$$

$$\frac{\tilde{v}_r^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \tilde{\Psi}_*^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \frac{\eta_r^i}{\theta^i} = C_r^i(\tilde{D}^i, q^i, r^i). \quad (35)$$

Given  $\tilde{\Lambda}_*^i$  and  $\tilde{\Lambda}_*^j$  in (15),  $\tilde{D}^i = D_*^i$  if both providers choose the first-best qualities. Therefore,

condition (34) is identical to (3), and condition (35) is identical to (4) simultaneously if and only if

$$\frac{1}{2\tau D_*^i} \left[ \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} + \tilde{\Psi}_*^i - C_D^i(D_*^i, q_*^i, r_*^i) \right] = \left[ \frac{v_r - \eta_r^i / \theta^i}{\tilde{v}_r^i} \right] = \left[ \frac{v_q - \eta_q^i / \theta^i}{\tilde{v}_q^i} \right]. \quad (36)$$

The last equality is necessary for first-best implementation, and is rewritten as (17) in the proposition.

Now suppose that (36) holds. The first equality in (36) can be rearranged to (18) in the proposition. Given the  $\tilde{\Psi}_*^i$  in (18),  $(q_*^i, r_*^i)$  is the unique solution to (34) and (35). Moreover, provider  $i$ 's equilibrium profit is zero if

$$\tilde{\Gamma}_*^i = -D_*^i \tilde{\Psi}_*^i + C^i(D_*^i, q_*^i, r_*^i). \quad (37)$$

Next, consider the special case in which  $\tilde{v}_k^i = v_k$  and  $\eta_q^i = \eta_r^i = 0$ . These equalities simplify the last two squared-bracket terms in (36) to 1, and hence, the first best is always implementable. Moreover, (36) can be simplified to

$$\begin{aligned} \tilde{\Psi}_*^i &= 2\tau D_*^i + C_D^i(D_*^i, q_*^i, r_*^i) \\ &= 2\tau \left[ \frac{1}{2} + \frac{v_q(q_*^i - q_*^j) + v_r(r_*^i - r_*^j) - \{C_D^i(D_*^i, q_*^i, r_*^i) - C_D^j(D_*^j, q_*^j, r_*^j)\}}{2\tau} \right] + C_D^i(D_*^i, q_*^i, r_*^i) \\ &= \tau + v_q(q_*^i - q_*^j) + v_r(r_*^i - r_*^j) + C_D^j(q_*^j, r_*^j, s_*^j), \end{aligned}$$

which is the optimal prospective price in Proposition 1.

**Proof of Proposition 4:** The first-order conditions of (19) with respect to  $q^i$  and  $r^i$  are

$$\frac{\tilde{v}_q^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \tilde{\Phi}_*^i + \tilde{\alpha}_*^i q^i + \tilde{\beta}_*^i r^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \left[ \frac{\eta_q^i}{\theta^i} + \tilde{\alpha}_*^i \right] = C_q^i(\tilde{D}^i, q^i, r^i) \quad (38)$$

$$\frac{\tilde{v}_r^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \tilde{\Phi}_*^i + \tilde{\alpha}_*^i q^i + \tilde{\beta}_*^i r^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \left[ \frac{\eta_r^i}{\theta^i} + \tilde{\beta}_*^i \right] = C_r^i(\tilde{D}^i, q^i, r^i). \quad (39)$$

Again, given  $\tilde{\Lambda}_*^i$  and  $\tilde{\Lambda}_*^j$  in (15),  $\tilde{D}^i = D_*^i$  if both providers choose the first-best qualities. Moreover, condition (38) is identical to (3), and (39) is identical to (4) if and only if

$$\begin{aligned} \tilde{\Phi}_*^i + \tilde{\alpha}_*^i q_*^i + \tilde{\beta}_*^i r_*^i - C_D^i(D_*^i, q_*^i, r_*^i) &= 2\tau D_*^i \frac{v_q - \eta_q^i / \theta^i - \tilde{\alpha}_*^i}{\tilde{v}_q^i} - \frac{\eta_q^i}{\theta^i} \\ \tilde{\Phi}_*^i + \tilde{\alpha}_*^i q_*^i + \tilde{\beta}_*^i r_*^i - C_D^i(D_*^i, q_*^i, r_*^i) &= 2\tau D_*^i \frac{v_r - \eta_r^i / \theta^i - \tilde{\beta}_*^i}{\tilde{v}_r^i} - \frac{\eta_r^i}{\theta^i}. \end{aligned}$$

And provider  $i$ 's equilibrium profit is zero if and only if

$$\tilde{\Phi}_*^i + \tilde{\alpha}_*^i q_*^i + \tilde{\beta}_*^i r_*^i = \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i}. \quad (40)$$

The three equations solve for

$$\begin{aligned}\tilde{\Phi}_*^i &= \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} + (\tilde{\varepsilon}_{q_*}^i + \tilde{\varepsilon}_{r_*}^i) \left[ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - C_D^i(D_*^i, q_*^i, r_*^i) + \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \right] \\ &\quad - \left[ v_q - \frac{\eta_q^i}{\theta^i} \right] q_*^i - \left[ v_r - \frac{\eta_r^i}{\theta^i} \right] r_*^i,\end{aligned}\quad (41)$$

$\tilde{\alpha}_*^i$  satisfying (20), and  $\tilde{\beta}_*^i$  satisfying (21). Finally, by setting  $\tilde{v}_k^i = v_k$  and  $\eta_q^i = \eta_r^i = 0$ , (20) and (21) become (13) and (14), and (41) can be simplified to

$$\Phi_*^i = \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} + (\varepsilon_{q_*}^i + \varepsilon_{r_*}^i) \left[ \frac{C^i(D_*^i, q_*^i, r_*^i)}{D_*^i} - C_D^i(D_*^i, q_*^i, r_*^i) \right] - v_q q_*^i - v_r r_*^i \quad (42)$$

in Proposition 2.

**Proof of Proposition 5:** The first-order conditions of (22) with respect to  $q^i, r^i$  are

$$\frac{\tilde{v}_q^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \hat{\Phi}_*^i + \hat{\alpha}_*^i q^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \left[ \frac{\eta_q^i}{\theta^i} + \tilde{\alpha}_*^i \right] = C_q^i(\tilde{D}^i, q^i, r^i) \quad (43)$$

$$\frac{\tilde{v}_r^i}{2\tau} \left[ \frac{\eta_q^i q^i + \eta_r^i r^i}{\theta^i} + \hat{\Phi}_*^i + \hat{\alpha}_*^i q^i - C_D^i(\tilde{D}^i, q^i, r^i) \right] + \tilde{D}^i \frac{\eta_r^i}{\theta^i} = C_r^i(\tilde{D}^i, q^i, r^i). \quad (44)$$

Again, given  $\tilde{\Lambda}_*^i$  and  $\tilde{\Lambda}^j$  in Lemma 4,  $\tilde{D}^i = D_*^i$  if both providers choose the first-best qualities. In addition, conditions (43) and (44) are, respectively, identical to (3) and (4), and provider  $i$ 's profit is zero by choosing  $(q_*^i, r_*^i)$  if and only if

$$\begin{aligned}\hat{\Phi}_*^i + \hat{\alpha}_*^i q_*^i - C_D^i(D_*^i, q_*^i, r_*^i) &= 2\tau D_*^i \frac{v_q - \eta_q^i / \theta^i - \hat{\alpha}_*^i}{\tilde{v}_q^i} - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \\ \hat{\Phi}_*^i + \hat{\alpha}_*^i q_*^i - C_D^i(D_*^i, q_*^i, r_*^i) &= 2\tau D_*^i \frac{v_r - \eta_r^i / \theta^i}{\tilde{v}_r^i} - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} \\ \hat{\Phi}_*^i + \hat{\alpha}_*^i q_*^i &= \frac{C^i(D_*^i, q_*^i, r_*^i) - \hat{\Gamma}_*^i}{D_*^i}.\end{aligned}\quad (45)$$

The three equations solve for  $\hat{\alpha}_*^i$  satisfying (23), and

$$\hat{\Phi}_*^i = \left[ 2\tau D_*^i \frac{v_r - \eta_r^i / \theta^i}{\tilde{v}_r^i} - \frac{\eta_q^i q_*^i + \eta_r^i r_*^i}{\theta^i} + C_D^i(D_*^i, q_*^i, r_*^i) \right] - \hat{\alpha}_*^i q_*^i$$

where the squared-bracket term is the value of  $\tilde{\Psi}_*^i$  in (18), and

$$\hat{\Gamma}_*^i = -D_*^i \left[ \hat{\Phi}_*^i + \hat{\alpha}_*^i q_*^i \right] + C^i(D_*^i, q_*^i, r_*^i) = -D_*^i \tilde{\Psi}_*^i + C^i(D_*^i, q_*^i, r_*^i)$$

$= \tilde{\Gamma}_*^i$  in (37). Finally, it is easy to check that  $\hat{\alpha}_*^i$  is strictly positive if and only if the left-hand side of (17) is larger than its right-hand side.

**Third best implementations and proof of Propositions 6 to 8:** In the third best, the social planner chooses uniform qualities  $q$  and  $r$  to maximize

$$v_q q + v_r r - \tau \left[ \int_0^{\tilde{D}^A} x dx + \int_{1-\tilde{D}^A}^1 (1-x) dx \right] - C^A(\tilde{D}^A, q, r) - C^B(\tilde{D}^B, q, r),$$

where  $\tilde{D}^i = \frac{1}{2} + \frac{(\tilde{v}_q^i - \tilde{v}_q^j)q + (\tilde{v}_r^i - \tilde{v}_r^j)r}{2\tau}$  is the misperceived consumers' demand for provider  $i$ 's service when the planner has no control over consumer allocation. The third-best qualities  $q_t$  and  $r_t$  are characterized by

$$\begin{aligned} v_q - \frac{\tilde{v}_q^A - \tilde{v}_q^B}{2\tau} \{ & [(\tilde{v}_q^A - \tilde{v}_q^B)q_t + (\tilde{v}_r^A - \tilde{v}_r^B)r_t] + [C_D^A(D_t^A, q_t, r_t) - C_D^B(D_t^B, q_t, r_t)] \} \\ & = C_q^A(D_t^A, q_t, r_t) + C_q^B(D_t^B, q_t, r_t) \end{aligned} \quad (46)$$

$$\begin{aligned} v_r - \frac{\tilde{v}_r^A - \tilde{v}_r^B}{2\tau} \{ & [(\tilde{v}_q^A - \tilde{v}_q^B)q_t + (\tilde{v}_r^A - \tilde{v}_r^B)r_t] + [C_D^A(D_t^A, q_t, r_t) - C_D^B(D_t^B, q_t, r_t)] \} \\ & = C_r^A(D_t^A, q_t, r_t) + C_r^B(D_t^B, q_t, r_t), \end{aligned} \quad (47)$$

where  $D_t^i = \frac{1}{2} + \frac{(\tilde{v}_q^i - \tilde{v}_q^j)q_t + (\tilde{v}_r^i - \tilde{v}_r^j)r_t}{2\tau}$  is the third-best demand for provider  $i$ 's service. Consider (46). When  $q$  increases,  $\{\tilde{v}_q^A - \tilde{v}_q^B\}/2\tau$  is the proportion of consumers who switch from provider  $B$  to provider  $A$  due to misconceptions. The first squared-bracket term measures the resulting change in total transportation cost. Similarly, the second squared-bracket term measures the resulting change in total service cost. When consumers have no misconceptions, or when consumer misconceptions are not provider-specific, so that  $\tilde{v}_q^i = \tilde{v}_q^j$  and  $\tilde{v}_r^i = \tilde{v}_r^j$ , (46) and (47) are, respectively, identical to (24) and (25). Thus, the second best is a special case of the third best.

Under PPS, provider  $i$ 's maximization program is the same as (19), and hence, the first-order conditions are the same as (34) and (39). These two conditions are identical to (46) and (47) simultaneously if and only if

$$\begin{aligned} & \frac{1}{2\tau} \left[ \frac{\eta_q^i q_t + \eta_r^i r_t}{\theta^i} + \tilde{\Psi}_t^i - C_D^i(D_t^i, q_t, r_t) \right] \\ & = \frac{1}{\tilde{v}_q^i} \left[ \{\text{Left-hand side of (46)}\} - C_q^j(D_t^j, q_t, r_t) - \frac{\eta_q^i}{\theta^i} \right] \\ & = \frac{1}{\tilde{v}_r^i} \left[ \{\text{Left-hand side of (47)}\} - C_r^j(D_t^j, q_t, r_t) - \frac{\eta_r^i}{\theta^i} \right]. \end{aligned}$$

In the second best in which  $\tilde{v}_k^i = v_k$  and  $\eta_q^i = \eta_r^i = 0$ , the above expression can be simplified to

$$\frac{1}{2\tau} \left[ \tilde{\Psi}_s^i - C_D^i(1/2, q_s, r_s) \right] = 1 - \frac{C_q^j(1/2, q_s, r_s)}{v_q} = 1 - \frac{C_r^j(1/2, q_s, r_s)}{v_r}.$$

The last equality is rewritten as (26) in Proposition 6.

Now consider P4P with both qualities contractible. Here, provider  $i$ 's maximization program is the same as (19) and the program's first-order conditions are the same as (38) and (39). The two conditions, together with the zero-profit condition (40) constitute a three-equation system. Setting the qualities in the three equations equal to  $q_t$  and  $r_t$  and rearranging terms yield the bonus rates and base price that implement the third best,

$$\begin{aligned}\tilde{\alpha}_t^i &= \left[ \frac{C_q^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_q^i}{\theta^i} \right] q_t - \tilde{\varepsilon}_{q_t}^i \left\{ \frac{C^i(D_t^i, q_t, r_t)}{D_t^i} - \left[ C_D^i(D_t^i, q_t, r_t) - \frac{\eta_q^i q_t + \eta_r^i r_t}{\theta^i} \right] \right\} \\ \tilde{\beta}_t^i &= \left[ \frac{C_r^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_r^i}{\theta^i} \right] r_t - \tilde{\varepsilon}_{r_t}^i \left\{ \frac{C^i(D_t^i, q_t, r_t)}{D_t^i} - \left[ C_D^i(D_t^i, q_t, r_t) - \frac{\eta_q^i q_t + \eta_r^i r_t}{\theta^i} \right] \right\} \\ \tilde{\Phi}_t^i &= \frac{C^i(D_t^i, q_t, r_t)}{D_t^i} + (\tilde{\varepsilon}_{q_t}^i + \tilde{\varepsilon}_{r_t}^i) \left\{ \frac{C^i(D_t^i, q_t, r_t)}{D_t^i} - \left[ C_D^i(D_t^i, q_t, r_t) - \frac{\eta_q^i q_t + \eta_r^i r_t}{\theta^i} \right] \right\} \\ &\quad - \left[ \frac{C_q^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_q^i}{\theta^i} \right] q_t - \left[ \frac{C_r^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_r^i}{\theta^i} \right] r_t,\end{aligned}$$

where  $\varepsilon_{k_t}^i = \frac{\tilde{v}_k^i k_t^i}{2\tau D_t^i}$ . In the second best in which  $\tilde{v}_k^i = v_k$  and  $\eta_q^i = \eta_r^i = 0$ , the optimal bonus rates and base price become (27) and (28) in Proposition 7 and

$$\begin{aligned}\Phi_s^i &= \frac{C^i(1/2, q_s, r_s)}{1/2} + (\tilde{\varepsilon}_{q_s} + \tilde{\varepsilon}_{r_s}) \left\{ \frac{C^i(1/2, q_s, r_s)}{1/2} - C_D^i(1/2, q_s, r_s) \right\} \\ &\quad - \frac{C_q^i(1/2, q_s, r_s)}{1/2} q_s - \frac{C_r^i(1/2, q_s, r_s)}{1/2} r_s.\end{aligned}$$

Finally, let quality  $r$  be non-contractible under P4P. In this set up, provider  $i$ 's maximization program is identical to (22) and the first-order conditions are identical to (43) and (44). At quality levels  $q_t$  and  $r_t$ , the two conditions, together with the zero profit condition (45) solve for

$$\begin{aligned}\hat{\alpha}_t^i &= \left[ \frac{C_q^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_q^i}{\theta^i} \right] - \frac{\tilde{v}_q^i}{\tilde{v}_r^i} \left[ \frac{C_r^i(D_t^i, q_t, r_t)}{D_t^i} - \frac{\eta_r^i}{\theta^i} \right] \\ \hat{\Phi}_t^i &= \left[ 2\tau \frac{C_r^i(D_t^i, q_t, r_t) - D_t^i \eta_r^i(q_t, r_t)/\theta^i}{\tilde{V}_r^i(q_t, r_t)} - \frac{\eta^i(q_t, r_t)}{\theta^i} + C_D^i(D_t^i, q_t, r_t) \right] - \hat{\alpha}_t^i q_t \\ \hat{\Gamma}_t^i &= -D_t^i \left[ \hat{\Phi}_t^i + \hat{\alpha}_t^i q_t \right] + C^i(D_t^i, q_t, r_t).\end{aligned}$$

In second best in which  $\tilde{v}_k^i = v_k$  and  $\eta_q^i = \eta_r^i = 0$ , these expressions become (29) in Proposition 8 and

$$\begin{aligned}\hat{\Phi}_s^i &= \left[ 2\tau \frac{C_r^i(1/2, q_s, r_s)}{v_r} + C_D^i(1/2, q_s, r_s) \right] - \hat{\alpha}_s^i q_s \\ \hat{\Gamma}_s^i &= -1/2 \left[ \hat{\Phi}_s^i + \hat{\alpha}_s^i q_s \right] + C^i(1/2, q_s, r_s).\end{aligned}$$

## References

- Acemoglu, Daron, Michael Kremer, and Atif Mian. "Incentives in Markets, Providers, and Governments." *Journal of Law, Economics, and Organization* 24.2 (2008): 273-306.
- Armstrong, Mark. "Search and Ripoff Externalities." *Review of Industrial Organization* 47.3 (2015): 273-302.
- Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics* 130.4 (2015): 1623-1667.
- Bardey, David, Chiara Canta, and Jean-Marie Lozachmeur. "The Regulation of Health Care Providers' Payments when Horizontal and Vertical Differentiation Matter." *Journal of Health Economics* 31.5 (2012): 691-704.
- Bardey, David, Helmuth Cremer, and Jean-Marie Lozachmeur. "The Design of Insurance Coverage for Medical Products under Imperfect Competition." *Journal of Public Economics* 137 (2016): 28-37.
- Baron, David P., and Roger B. Myerson. "Regulating a Monopolist with Unknown Costs." *Econometrica* 50.4 (1982): 911-930.
- Beitia, Arantza. "Hospital Quality Choice and Market Structure in a Regulated Duopoly." *Journal of Health Economics* 22.6 (2003): 1011-1036.
- Benabou, Roland, and Jean Tirole. "Bonus Culture: Competitive Pay, Screening, and Multitasking." *Journal of Political Economy* 124.2 (2016): 305-370.
- Besley, Timothy, and Maitreesh Ghatak. "Competition and Incentives with Motivated Agents." *American Economic Review* 95.3 (2005): 616-636.
- Brekke, Kurt R., Tor Helge Holmas, and Odd Rune Straume. "Reference Pricing, Competition, and Pharmaceutical Expenditures: Theory and Evidence from a Natural Experiment." *Journal of Public Economics* 95.7 (2011): 624-638.
- Brekke, Kurt R., Robert Nuscheler, and Odd Rune Straume. "Quality and Location Choices under Price Regulation." *Journal of Economics & Management Strategy* 15.1 (2006): 207-227.
- Brekke, Kurt R., Roberto Cellini, Luigi Siciliani, and Odd Rune Straume. "Competition in Regulated Markets with Sluggish Beliefs about Quality." *Journal of Economics & Management Strategy* 21.1 (2012a): 131-178.
- Brekke, Kurt R., Luigi Siciliani, and Odd Rune Straume. "Quality Competition with Profit Constraints." *Journal of Economic Behavior & Organization* 84.2 (2012b): 642-659.
- Brekke, Kurt R., Luigi Siciliani, and Odd Rune Straume. "Hospital Mergers with Regulated Prices." *Scandinavian Journal of Economics* 119 (2017): 597-627.
- Chalkley, Martin, and James M. Malcomson. "Contracting for Health Services when Patient Demand does not Reflect Quality." *Journal of Health Economics* 17.1 (1998a): 1-19.

- Chalkley, Martin, and James M. Malcomson. "Contracting for Health Services with Unmonitored Quality." *Economic Journal* 108.449 (1998b): 1093-1110.
- Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. "The Price ain't Right? Hospital Prices and Health Spending on the Privately Insured." *National Bureau of Economic Research Working Paper* w21815, 2015.
- Doyle, Cathal, Laura Lennox, and Derek Bell. "A Systematic Review of Evidence on the Links between Patient Experience and Clinical Safety and Effectiveness." *BMJ Open*, 3.1 (2013): e001570.
- Eggleston, Karen. "Multitasking and Mixed Systems for Provider Payment." *Journal of Health Economics* 24.1 (2005): 211-223.
- Ellis, Randall P., and Thomas G. McGuire. "Optimal Payment Systems for Health Services." *Journal of Health Economics* 9.4 (1990): 375-396.
- Enthoven, Alain C. "The History and Principles of Managed Competition." *Health Affairs* 12 suppl 1 (1993): 24-48.
- Farhi, Emmanuel, and Xavier Gabaix. "Optimal Taxation with Behavioral Agents." *National Bureau of Economic Research Working Paper* w21524, 2015.
- Finkelstein, Amy. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." *Quarterly Journal of Economics* 122.1 (2007): 1-37.
- Gaynor, Martin, Carol Propper, and Stephan Seiler. "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service." *American Economic Review* 106.11 (2016): 3521-57.
- Glaeser, Edward L., and Andrei Shleifer. "Not-for-Profit Entrepreneurs." *Journal of Public Economics* 81.1 (2001): 99-115.
- Gravelle, Hugh, Matt Sutton, and Ada Ma. "Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?" *Economic Journal* 120.542 (2010): F129-F156.
- Gutacker, Nils, Luigi Siciliani, Giuseppe Moscelli, and Hugh Gravelle. "Choice of Hospital: Which Type of Quality Matters?." *Journal of Health Economics* 50 (2016): 230-246.
- Hammond, Peter J. "Altruism." *The New Palgrave: A Dictionary of Economics* 1. Ed. John Eatwell, Murray Milgate and Peter Newman. London: Macmillan. 1987. 85-87.
- Hansmann, Henry. "The Ownership of Enterprise." Cambridge: Harvard University Press, 1996.
- Hart, Oliver, Andrei Shleifer, and Robert W. Vishny. "The Proper Scope of Government: Theory and an Application to Prisons." *Quarterly Journal of Economics* 112.4 (1997): 1127-1161.
- Holmstrom, Bengt, and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7 Special Issue (1991): 24-52.



- Kaarboe, Oddvar, and Luigi Siciliani. "Multi-Tasking, Quality and Pay for Performance." *Health Economics* 20.2 (2011): 225-238.
- Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112.2 (1997): 375-405.
- Kaiser Family Foundation and Health Services & Educational Trust (2016). Employer Health Benefits: 2016 Annual Survey, from <http://files.kff.org/attachment/Report-Employer-Health-Benefits-2016-Annual-Survey>
- Katz, Michael L. "Provider Competition and Healthcare Quality: More Bang for the Buck?" *International Journal of Industrial Organization* 31.5 (2013): 612-625.
- Krinsky, Sam, Andrew M. Ryan, Tod Mijanovich, and Jan Blustein. "Variation in Payment Rates under Medicare's Inpatient Prospective Payment System." *Health Services Research* 52.2 (2016): 676-696.
- Kuhn, Michael, and Luigi Siciliani. "Performance Indicators for Quality with Costly Falsification." *Journal of Economics & Management Strategy* 18.4 (2009): 1137-1154.
- Ma, Ching-to Albert. "Health Care Payment Systems: Cost and Quality Incentives." *Journal of Economics & Management Strategy* 3.1 (1994): 93-112.
- Ma, Ching-to Albert, and James F. Burgess Jr. "Quality Competition, Welfare, and Regulation." *Journal of Economics* 58.2 (1993): 153-173.
- Ma, Ching-to Albert, and Thomas G. McGuire. "Optimal Health Insurance and Provider Payment." *American Economic Review* 87.4 (1997): 685-704.
- McGuire, Thomas G., and Michael H. Riordan. "Incomplete Information and Optimal Market Structure Public Purchases from Private Providers." *Journal of Public Economics* 56.1 (1995): 125-141.
- Milgrom, Paul. "Is Sympathy an Economic Value? Philosophy, Economics, and the Contingent Valuation Method." *Contingent Valuation. A Critical Assessment*. Ed. Jerry A. Hausman. Amsterdam: North-Holland. 1993. 417-442.
- Milstein, Ricarda, and Jonas Schreyoegg. "Pay for Performance in the Inpatient Sector: A Review of 34 P4P Programs in 14 OECD Countries." *Health Policy* 120.10 (2016): 1125-1140.
- Moscelli, Giuseppe, Luigi Siciliani, Nils Gutacker, and Hugh Gravelle. "Location, Quality and Choice of Hospital: Evidence from England 2002-2013." *Regional Science and Urban Economics* 60 (2016): 112-124.
- Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal. "Can You get What You Pay for? Pay-for-Performance and the Quality of Healthcare Providers." *Rand Journal of Economics* 41.1 (2010): 64-91.
- O'Donoghue, Ted, and Matthew Rabin. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review Papers and Proceedings* 93.2 (2003): 186-191.

- Schwartz, Aaron L., Bruce E. Landon, Adam G. Elshaug, Michael E. Chernew, and J. Michael McWilliams. "Measuring Low-Value Care in Medicare." *JAMA Internal Medicine* 174.7 (2014): 1067-1076.
- Sherry, Tisamarie B. "A Note on the Comparative Statics of Pay-for-Performance in Health Care." *Health Economics* 25.5 (2016): 637-644.
- Shleifer, Andrei. "A Theory of Yardstick Competition." *RAND Journal of Economics* 16.3 (1985): 319-327.
- Spence, A. Michael. "Monopoly, Quality, and Regulation." *Bell Journal of Economics* 6.2 (1975): 417-429.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press, 2008.
- Thomson, Sarah, Laura Schang, and Michael E. Chernew. "Value-Based Cost Sharing in the United States and Elsewhere can Increase Patients' Use of High-Value Goods and Services." *Health Affairs* 32.4 (2013): 704-712.
- VanLare, Jordan M., and Patrick H. Conway. "Value-Based Purchasing—National Programs to Move from Volume to Value." *New England Journal of Medicine* 367.4 (2012): 292-295.
- White, Chapin, and Megan Eguchi. "Reference Pricing: A Small Piece of the Health Care Price and Quality Puzzle." *National Institute for Health Care Reform Research Brief* 18, 2014.
- Wu, Yaping, Yijuan Chen, and Sanxi Li. "Optimal Compensation Rule under Provider Adverse Selection and Moral Hazard." *Health Economics* (forthcoming).
- Wolinsky, Asher. "Regulation of Duopoly: Managed Competition vs Regulated Monopolies." *Journal of Economics & Management Strategy* 6.4 (1997): 821-847.

## Appendix B

### The comparative statics of $\tau$ in the first best

Without loss of generality, suppose that provider  $A$  is more efficient, so that in the first best  $D_*^A > 1/2$ . According to Theorems 5 and 6 in Milgrom and Shannon (1994), the solutions,  $-q_*^A$ ,  $-r_*^A$ ,  $q_*^B$ ,  $r_*^B$ , and  $-D_*^A$ , that maximize (1) are monotone increasing in  $\tau$  when all the pairwise cross-partial derivatives of  $-q_*^A$ ,  $-r_*^A$ ,  $q_*^B$ ,  $r_*^B$ ,  $-D_*^A$ , and  $\tau$  are nonnegative. These pairwise cross-partial derivatives are

$$\begin{aligned}
-\frac{\partial^2 W}{\partial \tau \partial q^A} &= 0 & -\frac{\partial^2 W}{\partial \tau \partial r^A} &= 0 & \frac{\partial^2 W}{\partial \tau \partial q^B} &= 0 \\
\frac{\partial^2 W}{\partial \tau \partial r^B} &= 0 & -\frac{\partial^2 W}{\partial \tau \partial D^A} &= (2D^A - 1) & -\frac{\partial^2 W}{\partial q^A \partial q^B} &= 0 \\
\frac{\partial^2 W}{\partial q^A \partial r^A} &= -C_{qr}^A & -\frac{\partial^2 W}{\partial q^A \partial r^B} &= 0 & \frac{\partial^2 W}{\partial q^A \partial D^A} &= v_q - C_{Dq}^A \\
-\frac{\partial^2 W}{\partial q^B \partial r^A} &= 0 & \frac{\partial^2 W}{\partial q^B \partial r^B} &= -C_{qr}^B & -\frac{\partial^2 W}{\partial q^B \partial D^A} &= v_q - C_{Dq}^B \\
-\frac{\partial^2 W}{\partial r^A \partial r^B} &= 0 & \frac{\partial^2 W}{\partial r^A \partial D^A} &= v_r - C_{Dr}^A & -\frac{\partial^2 W}{\partial r^B \partial D^A} &= v_r - C_{Dr}^B,
\end{aligned}$$

where the function  $C^i(D^i, q^i, r^i)$  is rewritten as  $C^i$  to ease notation. Because  $D_*^A > 1/2$ , the necessary conditions for  $-q_*^A$ ,  $-r_*^A$ ,  $q_*^B$ ,  $r_*^B$ ,  $-D_*^A$  to be monotone increasing in  $\tau$  are

$$-C_{qr}^A > 0, \quad -C_{qr}^B > 0, \quad v_q - C_{Dq}^A > 0, \quad v_q - C_{Dq}^B > 0, \quad v_r - C_{Dr}^A > 0, \quad v_r - C_{Dr}^B > 0.$$

Using the same method, it can be shown that  $q_*^A$ ,  $r_*^A$ ,  $q_*^B$ ,  $r_*^B$ ,  $D_*^A$  are monotone decreasing in  $\tau$  if

$$-C_{qr}^A > 0, \quad -C_{qr}^B > 0, \quad v_q - C_{Dq}^A > 0, \quad v_q - C_{Dq}^B < 0, \quad v_r - C_{Dr}^A > 0, \quad v_r - C_{Dr}^B < 0.$$

### Equilibrium existence

**Sufficient conditions for (5) to be quasi-concave and the corresponding best response map to be a contraction:** Let  $\Pi^i$  denotes profit function (5), and subscripts denote partial derivatives.  $\Pi^i$  is quasi-concave if

$$\begin{bmatrix}
0 & \Pi_{q^i}^i & \Pi_{r^i}^i \\
\Pi_{q^i}^i & \Pi_{q^i q^i}^i & \Pi_{q^i r^i}^i \\
\Pi_{r^i}^i & \Pi_{r^i q^i}^i & \Pi_{r^i r^i}^i
\end{bmatrix} \geq 0, \tag{48}$$

and

$$\begin{bmatrix} 0 & \Pi_{q^i}^i & \Pi_{r^i}^i & \Pi_{\Lambda^i}^i \\ \Pi_{q^i}^i & \Pi_{q^i q^i}^i & \Pi_{q^i r^i}^i & \Pi_{q^i \Lambda^i}^i \\ \Pi_{r^i}^i & \Pi_{r^i q^i}^i & \Pi_{r^i r^i}^i & \Pi_{r^i \Lambda^i}^i \\ \Pi_{\Lambda^i}^i & \Pi_{\Lambda^i q^i}^i & \Pi_{\Lambda^i r^i}^i & \Pi_{\Lambda^i \Lambda^i}^i \end{bmatrix} \leq 0 \quad (49)$$

(Takayama, 1985, p. 123). The sufficient conditions in terms of  $\tau$ ,  $v_q$ ,  $v_r$ , and  $C^i$  are tedious but uninformative, and omitted.

Using the same notation, the best response map of provider  $i$  satisfies the dominant diagonal condition, and hence, is a contraction if

$$-\Pi_{q^i q^i}^i > \Pi_{q^i r^i}^i + \Pi_{q^i \Lambda^i}^i + \Pi_{q^i q^j}^i + \Pi_{q^i r^j}^i + \Pi_{q^i \Lambda^j}^i \quad (50)$$

$$-\Pi_{r^i r^i}^i > \Pi_{r^i q^i}^i + \Pi_{r^i \Lambda^i}^i + \Pi_{r^i q^j}^i + \Pi_{r^i r^j}^i + \Pi_{r^i \Lambda^j}^i \quad (51)$$

$$-\Pi_{\Lambda^i \Lambda^i}^i > \Pi_{\Lambda^i q^i}^i + \Pi_{\Lambda^i r^i}^i + \Pi_{\Lambda^i q^j}^i + \Pi_{\Lambda^i r^j}^i + \Pi_{\Lambda^i \Lambda^j}^i \quad (52)$$

(Milgrom and Robert, 1990, p. 1271). These conditions guarantee that the game has a unique pure strategy Nash equilibrium.

**Sufficient conditions for the managed competition game to have a unique pure strategy Nash equilibrium in the subgame in Stage 2:** Continue to use  $\Pi^i$  to denote the profit function of provider  $i$ . It is sufficient to make sure that  $\Pi^i$  is quasi-concave and the best response map of provider  $i$  is a contraction. Since  $\Lambda^i$  is set by the payer,  $\Pi^i$  is quasi-concave as long as condition (48) holds. Moreover, the corresponding best-response map of provider  $i$  is a contraction if

$$-\Pi_{q^i q^i}^i > \Pi_{q^i r^i}^i + \Pi_{q^i q^j}^i + \Pi_{q^i r^j}^i \quad (53)$$

$$-\Pi_{r^i r^i}^i > \Pi_{r^i q^i}^i + \Pi_{r^i q^j}^i + \Pi_{r^i r^j}^i. \quad (54)$$

I assume the sufficient conditions hold throughout the rest of the paper.

## Cost-reduction effort and prospective payment system

Suppose that in addition to qualities, providers also choose cost-reduction efforts. To ease notation, let quality be unidimensional and be denoted by  $q^i$ . Following Hart et al. (1997), let the cost-reduction effort  $e^i$  lower provider  $i$ 's service cost and its consumers' service benefit. Both quality and cost-reduction efforts generate disutilities to the providers. More specifically, a consumer's benefit from receiving service with quality  $q^i$  and cost effort  $e^i$  is  $v(q^i - e^i)$ . Provider  $i$ 's service cost is  $C^i(D^i, q^i, e^i)$ , where  $C^i$  is twice differentiable,  $C_D^i > 0$ ,  $C_q^i > 0$ ,  $C_e^i < 0$ , strictly positive, and strictly convex in quality and cost effort. Finally, provider  $i$ 's total disutility of quality and cost effort is  $F^i(q^i, e^i)$ , where  $F^i$  is twice differentiable,  $F_q^i > 0$ ,  $F_e^i > 0$ , and is strictly convex.

Observe that this specification is more general than Ma (1994), who assumes that cost-reduction effort lowers cost, but not quality.

In this specification, social welfare is

$$\int_0^{D^A} [v(q^A - e^A) - \tau x] dx - C^A(D^A, q^A, e^A) - F^A(q^A, e^A) \\ + \int_{1-D^A}^1 [v(q^B - e^B) - \tau(1-x)] dx - C^B(D^B, q^B, e^B) - F^B(q^A, e^A),$$

and the first best is defined by

$$\frac{1}{2} + \frac{v(q_*^i - q_*^j - e_*^i + e_*^j) - \{C_D^i(D_*^i, q_*^i, e_*^i) - C_D^j(D_*^j, q_*^j, e_*^j)\}}{2\tau} = D_*^i \quad (55)$$

$$D_*^i v = C_q^i(D_*^i, q_*^i, e_*^i) + F_q^i(q_*^i, e_*^i) \quad (56)$$

$$-D_*^i v = C_e^i(D_*^i, q_*^i, e_*^i) + F_e^i(q_*^i, e_*^i). \quad (57)$$

Let  $-C_e^i(D_*^i, q_*^i, e_*^i) - D_*^i v > 0$ , so that cost-reduction effort is socially beneficial in the first best.

Under PPS, the maximization program of provider  $i$  is

$$\max_{q^i, e^i} \left[ \frac{1}{2} + \frac{v(q^i - q^j - e^i + e^j) - \{\Lambda_*^i - \Lambda_*^j\}}{2\tau} \right] \Psi_*^i - C^i(D^i, q^i, e^i) - F^i(q^i, e^i) + \Gamma_*^i. \quad (58)$$

To implement the first-best consumer allocation, it is necessary to set

$$\Lambda_*^i - \Lambda_*^j = C_D^i(D_*^i, q_*^i, e_*^i) - C_D^j(D_*^j, q_*^j, e_*^j).$$

Next, differentiating (58) with respect to  $q^i, e^i$  and setting the derivatives to zero, I get

$$\frac{v}{2\tau} [\Psi_*^i - C_D^i(D^i, q^i, e^i)] = C_q^i(D^i, q^i, e^i) + F_q^i(q^i, e^i) \quad (59)$$

$$-\frac{v}{2\tau} [\Psi_*^i - C_D^i(D^i, q^i, e^i)] = C_e^i(D^i, q^i, e^i) + F_e^i(q^i, e^i). \quad (60)$$

Conditions (59) and (60) are, respectively, identical to (56) and (57) if and only if

$$D_*^i v = \frac{v}{2\tau} [\Psi_*^i - C_D^i(D_*^i, q_*^i, e_*^i)] \\ -D_*^i v = -\frac{v}{2\tau} [\Psi_*^i - C_D^i(D_*^i, q_*^i, e_*^i)].$$

Both equations can be simplified to

$$\Psi_*^i = 2\tau D_*^i + C_D^i(D_*^i, q_*^i, e_*^i) \\ = 2\tau \left[ \frac{1}{2} + \frac{v(q_*^i - q_*^j - e_*^i + e_*^j) - \{C_D^i(D_*^i, q_*^i, e_*^i) - C_D^j(D_*^j, q_*^j, e_*^j)\}}{2\tau} \right] + C_D^i(D_*^i, q_*^i, e_*^i) \\ = \tau + v\{(q_*^i - q_*^j) - (e_*^i - e_*^j)\} + C_D^j(q_*^j, r_*^j, e_*^j).$$

Given this  $\Psi_*^i$  and the optimal copayments, the  $(q_*^i, e_*^i)$  pair is the unique solution to (59) and (60), and  $D^i = D_*^i$ . Finally, provider  $i$ 's equilibrium profit is  $F^i(q_*^i, e_*^i)$  if

$$\Gamma_*^i = -D_*^i \Psi_*^i + C^i(D_*^i, q_*^i, e_*^i).$$

## Monopoly model

Suppose that only provider  $A$  at location 0 is active. Since there is only one provider, I will drop the superscript that denotes the provider's identity. The social welfare is

$$\int_0^D [v_q q + v_r r - \tau x] dx - C(D, q, r).$$

Further suppose that it is not optimal to serve the consumer at location 1. The first best is characterized by

$$\begin{aligned} \frac{v_q q_* + v_r r_* - C_D(D_*, q_*, r_*)}{\tau} &= D_* \\ D_* v_q &= C_q(D_*, q_*, r_*) \\ D_* v_r &= C_r(D_*, q_*, r_*) \end{aligned}$$

where  $D_* < 1$ .

It is well understood that an unregulated single-price monopoly does not maximize social surplus. Hence, consider the first-best implementation problems as in Section 4. The first-best consumer allocation is implementable only if

$$\begin{aligned} \frac{\tilde{v}_q q_* + \tilde{v}_r r_* - \tilde{\Lambda}_*}{\tau} &= \frac{v_q q_* + v_r r_* - C_D(D_*, q_*, r_*)}{\tau} \\ \tilde{\Lambda}_* &= C_D(D_*, q_*, r_*) + [(\tilde{v}_q - v_q)q_* - (\tilde{v}_r - v_r)r_*]. \end{aligned} \quad (61)$$

Under PPS, the maximization program of provider  $A$  is

$$\max_{q, r} \left[ \frac{\tilde{v}_q q + \tilde{v}_r r - \tilde{\Lambda}_*}{\tau} \right] \left[ \eta_q q + \eta_r r + \theta \tilde{\Psi}_* \right] - \theta C(\tilde{D}, q, r) + \theta \tilde{\Gamma}_*,$$

and the first-order conditions are

$$\begin{aligned} \frac{\tilde{v}_q}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \tilde{\Psi}_* - C_D(\tilde{D}, q, r) \right] + \tilde{D} \frac{\eta_q}{\theta} &= C_q(\tilde{D}, q, r) \\ \frac{\tilde{v}_r}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \tilde{\Psi}_* - C_D(\tilde{D}, q, r) \right] + \tilde{D} \frac{\eta_r}{\theta} &= C_r(\tilde{D}, q, r). \end{aligned}$$

Given the value of  $\tilde{\Lambda}_*$  in (61), the first best is implementable if and only if

$$\frac{1}{\tau D_*} \left[ \frac{\eta_q q_* + \eta_r r_*}{\theta} + \tilde{\Psi}_* - C_D(D_*, q_*, r_*) \right] = \frac{v_r - \eta_r/\theta}{\tilde{v}_r} = \frac{v_q - \eta_q/\theta}{\tilde{v}_q}. \quad (62)$$

And if (62) is satisfied, the optimal  $\tilde{\Psi}_*$  and  $\tilde{\Gamma}_*$  are

$$\begin{aligned} \tilde{\Psi}_* &= \left[ \tau D_* \left( \frac{v_r - \eta_r/\theta}{\tilde{v}_r} \right) - \frac{\eta_q q_* + \eta_r r_*}{\theta} \right] + C_D(D_*, q_*, r_*) \\ \tilde{\Gamma}_* &= -D_* \tilde{\Psi}_* + C(D_*, q_*, r_*). \end{aligned}$$

When  $\eta_q = \eta_r = 0$  and  $\tilde{v}_k = v_k$ , (62) is always satisfied. Moreover, the consumer cost share and prospective price become  $\Lambda_* = C_D(D_*, q_*, r_*)$  and  $\Psi_* = \tau D_* + C_D(D_*, q_*, r_*)$ .

Under P4P, when both qualities are contractible, the maximization program of provider  $A$  is

$$\max_{q,r} \left[ \frac{\tilde{v}_q q + \tilde{v}_r r - \tilde{\Lambda}_*}{\tau} \right] \left[ \eta_q q + \eta_r r + \theta \left\{ \tilde{\Phi}_* + \tilde{\alpha}_* q + \tilde{\beta}_* r \right\} \right] - \theta C(\tilde{D}, q, r) + \theta \tilde{\Gamma}_*,$$

and the first-order conditions are

$$\begin{aligned} \frac{\tilde{v}_q}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \tilde{\Phi}_* + \tilde{\alpha}_* q + \tilde{\beta}_* r - C_D(\tilde{D}, q, r) \right] + \tilde{D} \left[ \frac{\eta_q}{\theta} + \tilde{\alpha}_* \right] &= C_q(\tilde{D}, q, r) \\ \frac{\tilde{v}_r}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \tilde{\Phi}_* + \tilde{\alpha}_* q + \tilde{\beta}_* r - C_D(\tilde{D}, q, r) \right] + \tilde{D} \left[ \frac{\eta_r}{\theta} + \tilde{\beta}_* \right] &= C_r(\tilde{D}, q, r). \end{aligned}$$

Given the value of  $\tilde{\Lambda}_*$  in (61), provider  $A$  chooses the first-best qualities and makes zero profit in equilibrium if and only if

$$\begin{aligned} \tilde{\Phi}_* + \tilde{\alpha}_* q_* + \tilde{\beta}_* r_* - C_D(D_*, q_*, r_*) &= 2\tau D_* \frac{v_q - \eta_q/\theta - \tilde{\alpha}_*}{\tilde{v}_q} - \frac{\eta_q q_* + \eta_r r_*}{\theta} \\ \tilde{\Phi}_* + \tilde{\alpha}_* q_* + \tilde{\beta}_* r_* - C_D(D_*, q_*, r_*) &= 2\tau D_* \frac{v_r - \eta_r/\theta - \tilde{\beta}_*}{\tilde{v}_r(q_*, r_*)} - \frac{\eta_q q_* + \eta_r r_*}{\theta} \\ \tilde{\Phi}_* + \tilde{\alpha}_* q_* + \tilde{\beta}_* r_* &= \frac{C(D_*, q_*, r_*)}{D_*}. \end{aligned}$$

The three-equation system solves for

$$\begin{aligned} \tilde{\alpha}_* q_* &= \left[ v_q - \frac{\eta_q}{\theta} \right] q_* - \tilde{\varepsilon}_{q_*} \left\{ \frac{C(D_*, q_*, r_*)}{D_*} - \left[ C_D(D_*, q_*, r_*) - \frac{\eta_q q_* + \eta_r r_*}{\theta} \right] \right\} \\ \tilde{\beta}_* r_* &= \left[ v_r - \frac{\eta_r}{\theta} \right] r_* - \tilde{\varepsilon}_{r_*} \left\{ \frac{C(D_*, q_*, r_*)}{D_*} - \left[ C_D(D_*, q_*, r_*) - \frac{\eta_q q_* + \eta_r r_*}{\theta} \right] \right\} \\ \tilde{\Phi}_* &= \frac{C(D_*, q_*, r_*)}{D_*} + (\tilde{\varepsilon}_{q_*} + \tilde{\varepsilon}_{r_*}) \left\{ \frac{C(D_*, q_*, r_*)}{D_*} - \left[ C_D(D_*, q_*, r_*) - \frac{\eta_q q_* + \eta_r r_*}{\theta} \right] \right\} \\ &\quad - \left[ v_q - \frac{\eta_q}{\theta} \right] q_* - \left[ v_r - \frac{\eta_r}{\theta} \right] r_*, \end{aligned}$$

where  $\tilde{\varepsilon}_{k_*} = \frac{v_k}{\tau} \frac{k_*}{D_*}$  is again the quality elasticity of demand at  $k_*$ . Again, when  $\eta_q = \eta_r = 0$  and  $\tilde{v}_k = v_k$ , the optimal bonuses become

$$\begin{aligned} \alpha_* q_* &= v_q q_* - \varepsilon_{q_*} \left\{ \frac{C(D_*, q_*, r_*)}{D_*} - C_D(D_*, q_*, r_*) \right\} \\ \beta_* r_* &= v_r r_* - \varepsilon_{r_*} \left\{ \frac{C(D_*, q_*, r_*)}{D_*} - C_D(D_*, q_*, r_*) \right\}. \end{aligned}$$

Next, consider P4P when only quality  $q$  is contractible. Provider  $A$ 's maximization program is

$$\max_{q,r} \left[ \frac{\tilde{v}_q q + \tilde{v}_r r - \tilde{\Lambda}_*}{\tau} \right] \left[ \eta_q q + \eta_r r + \theta \left\{ \hat{\Phi}_* + \hat{\alpha}_* q \right\} \right] - \theta C(\tilde{D}, q, r) + \theta \hat{\Gamma}_*,$$

and the first-order conditions are

$$\begin{aligned}\frac{\tilde{v}_q}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \hat{\Phi}_* + \hat{\alpha}_* q - C_D(\tilde{D}, q, r) \right] + \tilde{D} \left[ \frac{\eta_q}{\theta} + \tilde{\alpha}_* \right] &= C_q(\tilde{D}, q, r) \\ \frac{\tilde{v}_r}{\tau} \left[ \frac{\eta_q q + \eta_r r}{\theta} + \hat{\Phi}_* + \hat{\alpha}_* q - C_D(\tilde{D}, q, r) \right] + \tilde{D} \frac{\eta_r}{\theta} &= C_r(\tilde{D}, q, r).\end{aligned}$$

Given the value of  $\tilde{\Lambda}_*$  in (61), the first best is implemented at zero provider profit if and only if

$$\begin{aligned}\hat{\Phi}_* + \hat{\alpha}_* q_* - C_D(D_*, q_*, r_*) &= \tau D_* \frac{v_q - \eta_q/\theta - \hat{\alpha}_*}{\tilde{v}_q} - \frac{\eta_q q_* + \eta_r r_*}{\theta} \\ \hat{\Phi}_* + \hat{\alpha}_* q_* - C_D(D_*, q_*, r_*) &= \tau D_* \frac{v_r - \eta_r/\theta}{\tilde{v}_r} - \frac{\eta_q q_* + \eta_r r_*}{\theta} \\ \hat{\Phi}_* + \hat{\alpha}_* q_* &= \frac{C(D_*, q_*, r_*) - \hat{\Gamma}_*}{D_*}.\end{aligned}$$

The three equations solve for

$$\begin{aligned}\hat{\alpha}_*^i &= \left[ v_q - \frac{\eta_q}{\theta^i} \right] - \frac{\tilde{v}_q}{\tilde{v}_r} \left[ v_r - \frac{\eta_r}{\theta} \right] \\ \hat{\Phi}_* &= \left[ \tau D_* \frac{v_r - \eta_r/\theta}{\tilde{v}_r} - \frac{\eta_q q_* + \eta_r r_*}{\theta} + C_D(D_*, q_*, r_*) \right] - \hat{\alpha}_* q_* \\ \hat{\Gamma}_* &= -D_* \left[ \hat{\Phi}_* + \hat{\alpha}_* q_* \right] + C(D_*, q_*, r_*).\end{aligned}$$

## Reference

- Milgrom, Paul, and John Roberts. "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities." *Econometrica* 58.6 (1990): 1255-1277.
- Milgrom, Paul, and Chris Shannon. "Monotone Comparative Statics." *Econometrica* 62.1 (1994): 157-180.
- Takayama, Akira. *Mathematical Economics*. Cambridge: Cambridge University Press, 1985.