Designs of Tiered and Value-based Health Plans

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

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

March 2013

Keywords: tier, value, information, pricing, payment, health plan, incentive, quality, cost

1 Extended Abstract

Provider tiering and value-based health care networks are increasing popular, especially in the United States. For example, since 2011, Massachusetts requires all health plans in the state to offer at least one tiered network. A health plan grades providers in a tiered network on the basis of service quality and cost efficiency. The health plan then designates the providers to different tiers without disclosing providers' quality and cost information to consumers. Consumers pay different copayments to obtain services form providers in different tiers. For example, a consumer have to pay a $10 copayment to visit a provider in the Excellent tier, but a $20 copayment to visit a provider in the Standard tier.

Tiered and value-based networks represent the latest innovations in health care markets. However, they also raise some questions that cannot be addressed by the health economics literature. For example, how should a health plan grades its providers? How should a health plan determines copayments in different tiers? How should a health plan contracts with providers? How will consumers and providers respond to network design? In this paper, we address these problems in a model where oligopolistic health care providers complete for consumers. We show that a health care intermediary can achieve socially optimal service quality, cost efficiency, and consumption in tiered and value-based networks, even when none of these choices is contractible. Our model demonstrates the interdependence of consumer and provider behavior in health service markets, which is long ignored in the health economics literature. Our main results illustrate how consumer copayments, information design, and provider payments should be coordinated in an imperfectly competitive market.

In our model, profit-maximizing providers supply health services to consumers in a health plan. Service quality is costly, but a provider can invest in effort to reduce service cost. Providers differ in their efficiency. Moreover, neither quality nor cost effort is contractible. The health plan pays the providers by either cost reimbursement or prospective payment. On the other side, consumers' provider choices are also noncontractible. The covered consumers do not pay the full price of services. A consumer picks one of the horizontally differentiated providers based on service quality, copayment, and her horizontal preferences. Consumers do not directly observe providers' quality or cost effort choices, and have to obtain such information from the health plan.

We use the term Network to describe the health plan's information and copayment design. We study two forms of Network. In a Tiered Network, the health plan uses providers' quality and cost effort information to construct tiers. A provider belongs to a tier if its quality and effort are above certain thresholds. For example, a provider is qualify for Excellent tier only if its quality and cost effort are both above 85. The tier designation determines consumers' copayment for a provider's services.

We show how providers strategically respond to demand-side network design. We further show that the health plan can utilize providers' strategic responses to make efficient consumer choices and low copayments compatible. These results highlight the importance of studying consumer and provider incentives side-by-side. We provide market and technological conditions under which the health plan can achieve the socially optimum in a Tiered Network. We find that prospective payment outperforms cost reimbursement when premium services require high fixed cost and low variable cost. But the reverse is true when premium services require high fixed cost and high variable cost.

In a Value-based Network, the health plan uses providers' quality and cost information to construct value indexes. The health plan reports providers' value indexes to consumers. A provider's value index is the weighted sum of its quality and cost effort. For example, if a provider's quality is 90 and effort is 70, and the weight on quality is 0.6, the provider's value index is $82{=}90{\times}0.6{+}70{\times}0.4$. The health plan also sets the copayment for obtaining services from each provider.

We show that the health plan can always achieve the socially optimal service quality, cost efficiency, and consumption in a Value-based Network. In fact, social optimum can be achieve in a Value-based Network but not a Tiered Network if the fixed cost of premium services is low. However, Value-based Network gives providers more flexibility in quality and effort choices. It is also less effective than Tiered Network in stimulating demand responses to quality and effort. Provider payments are higher under both cost reimbursement and prospective payment in a Value-based Network.

2 The model

We now set up a model of health care network. A group of consumers will receive health services from one of two providers under a health plan. We discuss in turn these providers, the consumers, and the health plan structure.

2.1 Providers

Two providers, A and B, supply health services to consumers in a health plan. Each provider chooses a service quality and a cost-reduction effort. Let (q_i, e_i) denote the quality and cost effort, both nonnegative, chosen by Provider i, $i = A, B$. Under quality q and cost effort e, each provider's unit cost of serving a consumer is given by the function $C(q, e)$ for provider i. The unit cost function is strictly increasing in quality and strictly decreasing in effort ($C_q > 0 > C_e$), so higher quality means higher unit cost, but higher effort can reduce it. We assume that the function C is strictly convex, and also that the marginal unit cost of quality, $C_q(q, e)$, is nonincreasing in cost effort ($C_{qe}(q, e) \leq 0$).

While the two providers share the same variable cost structure, their fixed costs are different. One provider is more efficient than the other. A provider's fixed cost is determined by the quality and cost effect. If (q_A, e_A) denotes the quality and effort of Provider A, its fixed cost is $G(q_A) + H(e_A)$, where G and H are both strictly increasing and strictly convex functions. If (q_B, e_B) denotes the quality and effort of Provider B, its fixed cost is $\beta[G(q_B) + H(e_B)]$, with $\beta > 1$. Thus, the parameter β measures the inefficiency of Provider B relative to Provider A. Although Provider B is less efficient, it may still operate because some consumers may naturally prefer using its services, as we describe next.

2.2 Consumers

A set of consumers are covered by the health plan, and their total mass is normalized at 1. The consumers are described by their valuations of health services, v, and a proclivity parameter x towards the two providers. Each consumer has the same valuation, so v is a strictly positive constant and identical for all consumers. It is customary to use a Hotelling, horizontal product-differentiation structure to describe consumers' proclivity towards the two providers, so we let the consumers be uniformly distributed on the $[0, 1]$ interval. We use a nonnegative parameter τ to measure the strength of consumers' horizontal preferences. A consumer located at x incurs a mismatch disutility τx when obtaining services from Provider A, and a mismatch disutility $\tau(1-x)$ when obtaining services from Provider B. When τ is large, each provider faces a less elastic demand.

A consumer may have to pay a fee to use a provider. We will expand on fee determinations when we describe health plan networks in Subsection 2.4. Let s_A and s_B, both nonnegative, be the respective copayments when a consumer uses Provider A and Provider B. When health care qualities at Providers A and B are, respectively, q_A and q_B, a consumer at x obtains utilities $vq_A - s_A - \tau x$ and $vq_B - s_B - \tau(1 - x)$ from these providers. We let each consumer receive services from one of the two providers, hence the market is covered.

We assume that consumers have the same valuation on quality. Our interpretation is that we consider the set of consumers who view the quality of care of this service similarly. Within this set of consumers, the horizontal product differentiation aspect of consumer preferences can be broadly interpreted. The Hotelling line can literally mean distance, and in health care, travel costs are important determinants of demands. Alternatively, different providers may have different practice styles, even different hours of operations, as well as other support services. Consumers may have diverse preferences towards these attributes, which are captured by the location and intensity parameters, x and τ.

2.3 First best

The health plan writes contracts with providers and consumers. In the first best, providers' qualities and efforts choices, and consumers' provider choices are contractible. An allocation specifies each provider's quality and cost-reduction effort, as well as which provider should serve each consumer. The first-best allocation is one that maximizes aggregate consumer utilities less the production cost, which we call social welfare. It is obvious that the first best will assign consumers with small values of x to Provider A, and consumers with large values of x to Provider B. The health plan chooses these variables to maximize the sum of consumer utilities and profits. Let the health plan assigns the consumers with $x \leq \widehat{x}$ to Provider A, and the remaining consumers to Provider B. Social welfare is

$$\begin{aligned} W(q_A, q_B, e_A, e_B, \widehat{x}) \equiv & \int_0^{\widehat{x}} [vq_A - \tau x - C(q_A, e_A)]dx + \int_{\widehat{x}}^1 [vq_B - \tau(1-x) - C(q_B, e_B)]dx \\ & -[G(q_A) + H(e_A)] - \beta[G(q_B) + H(e_B)]. \end{aligned} \quad (1)$$

The first term in the welfare expression (1) is the sum of consumers' utilities less the variable costs for consumers obtaining services at Provider A; the second term is the corresponding value at Provider B. The last two terms in (1) are the total fixed costs of qualities and efforts. The following Lemma characterizes the first best.

Lemma 1 *: In the first best, the service qualities and cost-reduction efforts at the two providers, respectively,* q_A^*, q_B^*, e_A^*, e_B^*, *and consumer allocation across the two providers, namely* $\widehat{x}^*$, *are given by*

$$\widehat{x}^*[v - C_q(q_A^*, e_A^*)] = G'(q_A^*) \quad (2)$$

$$(1 - \widehat{x}^*)[v - C_q(q_B^*, e_B^*)] = \beta G'(q_B^*) \quad (3)$$

$$-\widehat{x}^* C_e(q_A^*, e_A^*) = H'(e_A^*) \quad (4)$$

$$-(1 - \widehat{x}^*) C_e(q_B^*, e_B^*) = \beta H'(e_B^*) \quad (5)$$

$$\frac{1}{2} + \frac{v(q_A^* - q_B^*) - [C(q_A^*, e_A^*) - C(q_B^*, e_B^*)]}{2\tau} = \widehat{x}^*. \quad (6)$$

Furthermore, $q_A^* > q_B^*$, $e_A^* > e_B^*$, *and* $\widehat{x}^* > 1/2$. *That is, the more efficient Provider A sets higher quality, more cost effort, and serves more consumers.*

Because Provider A is more efficient than Provider B (which has higher fixed costs ($\beta > 1$)), the first best prescribes that Provider A has higher quality, lower variable cost and serves more consumers. Equations (2) to (6) have the usual interpretations. Raising a provider's service quality increases consumer utilities, but also the provider's unit cost and fixed cost of quality. Equations (2) and (3) balance these marginal effects. Similarly, raising a provider's cost effort decreases unit cost, but increases fixed cost of effort. Equations (4) and (5) balance these two opposing effects. Equation (6) defines $\widehat{x}^*$, the consumer who receives the same net social benefit from either provider.

2.4 Health plan, payment, and information

We first describe the provider payment contracts. Because of the complexity of health care services, contracts that specify how providers choose qualities and cost-reduction efforts are infeasible. In practice, health plans do use financial (and renewal) contracts that are based on how much health care service and cost providers have supplied. Accordingly, we let qualities and cost-reduction efforts be noncontractible, but the quantities and unit costs of services $C(q_i, e_i)$, $i = A, B$, are *ex post* verifiable. We study two most common forms of provider payment. Under *prospective payment*, the health plan pays p_i to Provider i for a unit of services supplied. Under *cost reimbursement*, the health plan pays any *ex post* unit cost $C(q_i, e_i)$ plus a margin m_i to Provider i for a unit of services supplied. In this paper, we only consider these two forms of payments in the analytical models.[1]

Consumers' choices of providers are also noncontractible. As mentioned above, the health plan may impose copayments. Consumers do not observe directly providers' quality and cost effort choices; the health plan, however, does observe these choices. The health plan decides how to convey quality and cost-effort information to consumers. In the literature, consumers are assumed to observe providers' care quality. In our setup, this is equivalent to the health plan fully disclosing quality information to consumers. Here, we study a general information disclosure strategy.

In a *Tiered Network*, the health plan uses the information on quality and cost effort (q_i, e_i), $i = A, B$, gathered from the providers to construct tiers. We consider two tiers: *Excellent* and *Standard*. A provider belongs to the Excellent tier if its quality and cost effort are above some given thresholds; similarly for the Standard tier. (We define these thresholds formally in the next section.) After assessing providers' qualities and cost efforts, the health plan announces each provider's tier. Consumers' copayments for services obtained from a provider may depend on that provider's designated tier. For example, if Provider A is in the Excellent tier, then consumers may pay lower copayments when using it than if Provider A is in the Standard tier.

In a *Value-based Network*, after observing a provider's quality and cost effort, the health plan constructs

[1] Consumer cost heterogenity can be readily incorporated into this framework. See, for example, Ma (1994), Ma and Mak (2012).

a value index equal to a linearly weighted sum of the quality and the cost effort. (The value index will be formally defined in the section after next.) The health plan then discloses providers' quality indexes to consumers. The health plan also sets consumers' copayments for obtaining services at the two providers.

3 Tiered Network

3.1 Network structure and extensive form

We first lay out how the health plan constructs the tiers and copayments. Figure 1 illustrates the network structure. We use two triples, (q^{Ex}, e^{Ex}, s^{Ex}) and (q^{St}, e^{St}, s^{St}), to define health plans' tiers and copayment policies, where q, e, and s denote quality, cost effort, and consumer copayment. The health plan assigns Provider i to the Excellent tier if and only if quality and cost effort (q_i, e_i) are both higher than (q^{Ex}, e^{Ex}): $(q_i, e_i) \geq (q^{Ex}, e^{Ex}) \geq (0, 0)$. This is the upper-right region in Figure 1. Next, suppose that Provider i fails to qualify for the Excellent tier (because $q_i < q^{EX}$, $e < e^{EX}$, or both). Then, Provider i is assigned to the Standard tier if and only if quality and cost effort (q_i, e_i) are both higher than (q^{St}, e^{St}): $(q_i, e_i) \geq (q^{St}, e^{St}) \geq (0, 0)$, where $q^{St} < q^{Ex}$ and $e^{St} < e^{Ex}$. This is the shaded, L-shape region in Figure 1. A consumers pay $s^{Ex} \geq 0$ to obtain a unit of service from a provider in the Excellent tier, and $s^{St} \geq 0$ from one in the Standard tier. If a provider fails to achieve any tier, it is excluded from the network. (We will assume that insured consumers will not obtain service from an excluded provider.)

We study the following Tiered Network, extensive-form game:

Stage 1: The health plan sets (q^{Ex}, e^{Ex}, s^{Ex}) and (q^{St}, e^{St}, s^{St}). Under cost reimbursement, the health plan sets the margin m_i, $i = A, B$, and commits to reimbursing Provider i's operating cost. Under prospective payment, the health plan sets the price p_i, $i = A, B$, for each unit of service supplied by Provider i.

Stage 2: Providers A, and B choose quality and cost-reduction effort simultaneously.

Stage 3: The insurer observes the providers' chosen qualities and efforts, and reports the tier that each provider belongs to.

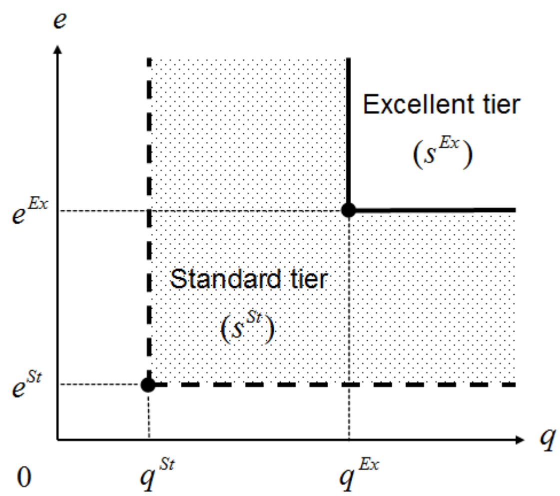

Figure 1: Tiers and copayments

Stage 4: Consumers learn the providers' designated tiers, and decide between obtaining services from Provider A and Provider B.

Consumers do not observe the providers' qualities and efforts, so, in Stage 4, must form beliefs about them knowing only the tier each provider belongs to. Therefore, we characterize prefect Bayesian equilibria of the Tired Network game. The health plan's objective is to implement the first best as a perfect Bayesian equilibrium. The next two subsections study, respectively, equilibria under cost reimbursement and prospective payment. Upon seeing providers' tier designation, consumers believe that the providers' qualities are q_A and q_B, and therefore evaluation their utilities from obtaining services there according to $vq_A - s_A - \tau x$ and $vq_B - s_B - \tau(1-x)$. Let $\widehat{x}$ be defined by $vq_A - s_A - \tau\widehat{x} = vq_B - s_B - \tau(1-\widehat{x})$, then market shares of Providers A and B are, respectively, $\widehat{x}_A = \widehat{x}$ and $\widehat{x}_B = 1-\widehat{x}$.

3.2 Cost reimbursement and first best

When consumers can directly observe qualities and cost efforts, a provider is free to choose any nonnegative pair of quality and cost effort to maximize profit. In the conventional setup, a big literature shows that cost reimbursement leads to zero cost effort and suboptimal quality. In the Tiered Network, consumers observe the providers' tiers but not their qualities and efforts. Our first result shows in any perfect-Bayesian equilibrium, an active provider's choice can only be either (q^{Ex}, e^{Ex}) or (q^{St}, e^{St}).

Lemma 2 : *Consider cost reimbursement. If a provider is assigned to the Excellent tier in a perfect-Bayesian equilibrium, its quality and cost effort are* (q^{Ex}, e^{Ex}). *Likewise, if a provider is assigned to the Standard tier in a perfect-Bayesian equilibrium, its quality and cost effort are* (q^{St}, e^{St}).

According to Lemma 2, in equilibrium a provider must use the lowest possible quality and cost effort to qualify for the Excellent or Standard tier. This is because in Stage 4 consumers only observe a provider's designated tier, but not its quality or effort choices. Consumers' beliefs on providers' qualities depend only on the tier designation, so the tier designation also determines each provider's market share and revenue. Among combinations of quality and cost effort that qualify a provider for a tier, a profit-maximizing provider must choose the least costly pair, one that minimizes $G(q) + H(e)$. Hence, in equilibrium, when consumers observe the providers' designated tiers, they must believe that the providers have chosen the lowest possible quality and cost effort in a qualified tier.

The health plan can utilize the tiers to incentivize optimal quality and cost effort choices. In fact, the following is an immediate consequence of Lemma 2.

Corollary 1 : *Consider cost reimbursement. The first best is implementable as a perfect-Bayesian equilibrium only if the health plan sets thresholds* $(q^{Ex}, e^{Ex}) = (q_A^*, e_A^*)$ *for the Excellent tier and* $(q^{St}, e^{St}) = (q_B^*, e_B^*)$ *for the Standard tier.*

We next derive the condition on the copayments for the implementation of the first best. Suppose that in equilibrium Providers A and B choose (q_A^*, e_A^*) and (q_B^*, e_B^*), respectively, in Stage 2. The health plan then in Stage 3 reports that Provider A belongs to the Excellent tier and Provider B to the Standard tier. By Lemma 2, consumers' demand for services at Provider A in Stage 4 is $1/2 + (1/2\tau)[v(q_A^* - q_B^*) - (s^{Ex} - s^{St})]$. To implement the first best, the health plan must choose s^{Ex} and s^{St} such that the consumer demand for provider A's services equals to $\widehat{x}^*$ in condition (6). Therefore, in a first-best equilibrium, the copayments must satisfy $s^{Ex} - s^{St} = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$: the difference between the copayments in the Excellent and Standard tiers must be equal to the difference in first-best unit costs of the more efficient Provider A and Provider B.

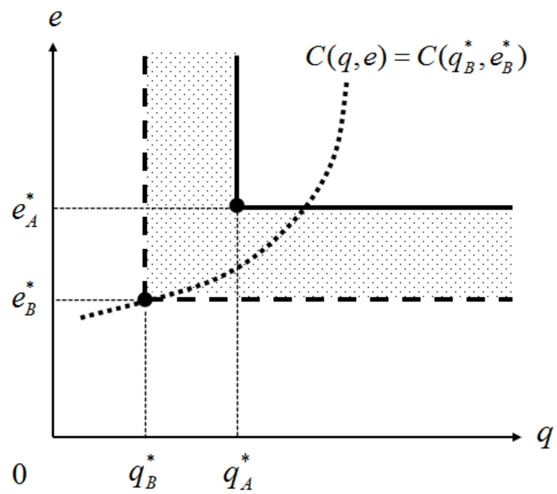

Figure 2: Equilibrium tiers and copayments

Lemma 3 *: Suppose the first best is implementable as a perfect-Bayesian equilibrium. The lowest, non-negative copayments are* $s^{Ex} = 0$, $s^{St} = C(q_B^*, e_B^*) - C(q_A^*, e_A^*)$ *if* $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$ *and* $s^{Ex} = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$, $s^{St} = 0$ *if* $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$.

Our assumptions do not pin down the relative magnitude between $C(q_A^*, e_A^*)$ and $C(q_B^*, e_B^*)$. Lemma 1 does prescribe higher quality and higher cost effort for Provider A, due to its lower fixed costs, so the operating cost $C(q_A^*, e_A^*)$ may be lower or higher than Provider B's operating cost $C(q_B^*, e_B^*)$. Consider a set of parameter configuration for which $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$. Define the iso-unit cost line by the combinations of q and e such that $C(q, e) = C(q_B^*, e_B^*)$. If $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$, then we have (q_A^*, e_A^*) located strictly above the iso-unit cost line. The dotted, upward-sloping curve in Figure 2 is the iso-unit cost curve of (the less efficient) Provider B at the first-best unit cost level.[2] In this case, the copayment for a Standard-tier provider is higher than an Excellent-tier provider, and the lowest nonnegative copayments are $s^{Ex} = 0$, $s^{St} = C(q_B^*, e_B^*) - C(q_A^*, e_A^*)$. Now if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$, then the first best requires $s^{Ex} = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$, and $s^{St} = 0$; the copayment at the Excellent tier is higher than Standard tier.

[2] The iso-unit cost curve is upward sloping because C is increasing in q and decreasing in e. Furthermore, the lower-contour set of a convex function is convex, and hence the iso-unit cost curve has the shaped as drawn.

Provider B

	tier	(q_A^*, e_A^*)	(q_B^*, e_B^*)	$(0,0)$
	(q_A^*, e_A^*)	$1/2, 1/2$	$\widehat{x}^*, 1-\widehat{x}^*$	$1, 0$
Provider A	(q_B^*, e_B^*)	$1-\widehat{x}^*, \widehat{x}^*$	$1/2, 1/2$	$1, 0$
	$(0,0)$	$0, 1$	$0, 1$	$0, 0$

Table 1: Equilibrium qualities, efforts, and market shares

We now study the providers' quality and effort choices. In Stage 2, the two providers simultaneously choose qualities and efforts to maximize profits. Because consumers' beliefs about qualities and efforts are constrained by Lemma 2, a profit-maximizing provider must choose (q^{Ex}, e^{Ex}) to qualify for the Excellent tier and (q^{St}, e^{St}) to qualify for the Standard tier. Let the tier thresholds and copayments be given by Corollay 1 and Lemma 3. Table 1 defines the providers' market shares in continuation games with different tier designations. For example, if Provider A chooses (q_A^*, e_A^*) and Provider B chooses (q_B^*, e_B^*), then Provider A qualifies for the Excellent tier, while Provider B qualifies for the Standard tier, and so the their market shares are $\widehat{x}^*$ and $1-\widehat{x}^*$, respectively. This shows up as the first row and second column in Table 1. If they both qualify for the Excellent tier or the Standard tier, they split the market equally; these are in the first-row-first-column, and second-row-second-cloumn cells in Table 1.

In a first-best equilibrium under cost reimbursement, Provider A must choose (q_A^*, e_A^*) and its profit is $m_A\widehat{x}^* - [G(q_A^*) + H(e_A^*)]$. Similarly, Provider B must choose (q_B^*, e_B^*) and its profit is $m_B(1-\widehat{x}^*) - \beta[G(q_B^*) + H(e_B^*)]$. To implement the first best, the health plan sets margins m_A and m_B such that neither provider finds unilateral devitation profitable.

The first best is implementable if there are margins m_A and m_B that satisfy the following incentive constraints:

$$m_A\widehat{x}^* - [G(q_A^*) + H(e_A^*)] \geq m_A/2 - [G(q_B^*) + H(e_B^*)] \quad (7)$$

$$m_B(1-\widehat{x}^*) - \beta[G(q_B^*) + H(e_B^*)] \geq m_B/2 - \beta[G(q_A^*) + H(e_A^*)] \quad (8)$$

and nonnegative profit constraints:

$$m_A\widehat{x}^* - [G(q_A^*) + H(e_A^*)] \geq 0 \quad (9)$$

$$m_B(1-\widehat{x}^*) - \beta[G(q_B^*) + H(e_B^*)] \geq 0. \quad (10)$$

In a first-best equilibrium, Provider A chooses (q_A^*, e_A^*). The incentive constraint (7) says that Provider A should not find deviating to (q_B^*, e_B^*) and capturing 1/2 of the market (see Table 1) profitable. The incentive constraint (8) has the same meaning. The incentive constraints guarantee that Provider A picking (q_A^*, e_A^*) and Provider B picking (q_B^*, e_B^*) are mutual best responses (given equilibrium consumer beliefs). The nonnegative profit constraints (9) and (10) guarantee that in equilibrium both providers prefer to be active. Can we find margins m_A and m_B to implement the first best as a perfect Bayesian equilibrium?

Proposition 1 : *Consider cost reimbursement. Let the tier and copayment policies be given by those in Corollay 1 and Lemma 3. The first best is implemented as a perfect-Baynesian equilibrium only if*

$$\frac{G(q_A^*) + H(e_A^*)}{G(q_B^*) + H(e_B^*)} \geq \frac{1}{2(1-\widehat{x}^*)}. \tag{11}$$

Proposition 1**,** together with Corollary 1 and Lemma 3, lay out the necessary conditions of implementing the first best as a perfect-Baynesian equilibrium. Proposition 1 says that the health plan can find margins m_A and m_B that satisfy constraints (7) to (10) only if condition (11) holds.

By Lemma 1, $\widehat{x}^* > 1/2$ and $G(q_A^*) + H(e_A^*) > G(q_B^*) + H(e_B^*)$. Therefore, both incentive constraint (7) and nonnegative constraint (9) require m_A to be sufficiently big. Only one of the two constraints is binding at a time: The nonnegative profit constraint (9) is slack when $\widehat{x}^*$ is small and $G(q_A^*) + H(e_A^*)$ is big relative to $G(q_B^*) + H(e_B^*)$. Here, deviating to (q_B^*, e_B^*) is more attractive to Provider A because the loss in revenue, $(\widehat{x}^* - 1/2)m_A$, is small but the cost saving, $G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)$, is big. The incentive constraint (7) is slack when $\widehat{x}^*$ is big and $G(q_A^*) + H(e_A^*)$ is small. In this case, deviating to (q_B^*, e_B^*) is less profitable and being inactive is more attractive to Provider A.

Conditions (8) and (10) constrain m_B from different directions. The incentive constraint (8) requires m_B to be sufficiently small, so that deviating to (q_A^*, e_A^*) and getting $(\widehat{x}^* - 1/2)$ more consumers is not profitable to Provider B. The nonnegative profit constraint (10) requires m_B to be sufficiently big, so that Provider B prefers to be active. Both constraints can be simultaneously satisfied only if condition (11) holds. That is, when $\widehat{x}^*$ is small and $G(q_A^*) + H(e_A^*)$ is big relative to $G(q_B^*) + H(e_B^*)$ and hence deviating to (q_A^*, e_A^*) is less profitable to Provider B.

Corollary 2 *The health plan minimizes the cost of implementing the first best by setting*

$$m_A = \frac{[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} \tag{12}$$

$$m_B = \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}. \tag{13}$$

Both conditions (8) and (9) are slack when the first best is implementable. To minimize the cost of implementing the first best, the health plan sets m_A and m_B such that constraints (7) and (10) are just binding. This gives the two expressions in the Proposition. In this first-best equilibrium, the less efficient Provider B makes zero profit because deviating to (q_A^*, e_A^*) results in negative profit, but the more efficient Provider A makes a positive profit because the provider has an incentive to undercut quality and effort to (q_B^*, e_B^*).

3.3 Prospective payment and first best

[To be added]

4 Value-based Network

4.1 Value indices and copayments

[To be added]

4.2 Cost reimbursement and first best

[To be added]

4.3 Prospective payment and first best

[To be added]

5 Optimal network

[To be added]

6 Conclusion

[To be added]

Appendix

Proof of Lemma 1: First, we obtain equations (2) to (6) by setting the first-order derivatives of (1) with respect to q_A, q_B, e_A, e_B, and $\widehat{x}$, to 0. These are necessary conditions for the maximization of (1).

We now show that $q_A^* > q_B^*$, $e_A^* > e_B^*$, and $\widehat{x}^* > 1/2$. By symmetry, if β was set at 1, then we would have $q_A^* = q_B^*$, $e_A^* = e_B^*$, $\widehat{x}^* = 1/2$. Therefore, it is sufficient to show that q_A^*, $-q_B^*$, e_A^*, $-e_B^*$, and $\widehat{x}^*$ are monotone increasing in β. Our assumptions on C and G and H have been laid out in Subsection 2.1. However, for the purpose here, we would strengthen them by restricting the sets of feasible quality and cost effort, q and e for each firm to those values where $v - C_q(q, e) > 0$. According to the first-order conditions (2) and (3), there cannot be any solution at which $v - C_q(q, e) < 0$. Such a restriction, $v - C_q(q, e) > 0$, does not affect the first best.

Now we can apply Theorems 5 and 6 in Milgrom and Shannon (1994), which say that solutions, q_A^*, $-q_B^*$, e_A^*, $-e_B^*$, and $\widehat{x}^*$, that maximize (1) are monotone increasing in β when all the pairwise cross-partial derivatives of q_A, $-q_B$, e_A, $-e_B$, $\widehat{x}$, and β are nonnegative. Indeed, we have:

$$\frac{\partial^2 W}{\partial\beta\partial q_A} = 0 \qquad -\frac{\partial^2 W}{\partial\beta\partial q_B} = G'(q_B) > 0 \qquad \frac{\partial^2 W}{\partial\beta\partial e_A} = 0$$

$$-\frac{\partial^2 W}{\partial\beta\partial e_B} = H'(e_B) > 0 \qquad \frac{\partial^2 W}{\partial\beta\partial\widehat{x}} = 0 \qquad -\frac{\partial^2 W}{\partial q_A\partial q_B} = 0$$

$$\frac{\partial^2 W}{\partial q_A\partial e_A} = -\widehat{x}C_{qe}(q_A, e_A) \geq 0 \qquad -\frac{\partial^2 W}{\partial q_A\partial e_B} = 0 \qquad \frac{\partial^2 W}{\partial q_A\partial\widehat{x}} = v - C_q(q_A, e_A) > 0$$

$$-\frac{\partial^2 W}{\partial q_B\partial e_A} = 0 \qquad \frac{\partial^2 W}{\partial q_B\partial e_B} = -(1-\widehat{x})C_{qe}(q_B, e_B) \geq 0 \qquad -\frac{\partial^2 W}{\partial q_B\partial\widehat{x}} = v - C_q(q_B, e_B) > 0$$

$$-\frac{\partial^2 W}{\partial e_A\partial e_B} = 0 \qquad \frac{\partial^2 W}{\partial e_A\partial\widehat{x}} = -Ce(q_A, e_A) > 0 \qquad -\frac{\partial^2 W}{\partial e_B\partial\widehat{x}} = -Ce(q_B, e_B) > 0.$$

■

Proof of Lemma 2: Suppose that the Lemma is false, so suppose that in a prefect-Baynesian equilibrium, Provider i is assigned to the Excellent tier and $(q^{Ex}, e^{Ex}) \neq (q_i, e_i)$. By the definition of the Excellent tier, we must have $q_i > q^{Ex}$, $e_i > e^{Ex}$, or both. Let the equilibrium margin and market share of provider i be m_i and $\widehat{x}_i$, respectively. Under cost reimbursement, Provider A's equilibrium profit is $m_i\widehat{x}_i - G(q_i) - H(e_i)$. Now let Provider i deviate from (q_i, e_i) to (q^{Ex}, e^{Ex}). Provider i still belongs to the Excellent tier, which means that consumers' beliefs about Provider A's quality remain the same, and so Provider A's market share must remain at $\widehat{x}_i$. But because G and H are strictly increasing, we have $m_i\widehat{x}_i - G(q^{Ex}) - H(e^{Ex}) > m_i\widehat{x}_i - G(q_i) - H(e_i)$, which says that the deviation is profitable, a contradiction. The proof for the Standard tier is similar, and omitted.■

Proof of Lemma 3: The Lemma follows immediately from the equation $s^{Ex} - s^{St} = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$ and the inequalities $s^{Ex} \geq 0$ and $s^{St} \geq 0$.■

Proof of Proposition 1: Rearranging inequalities (7) to (10), we have

$$m_A \geq \frac{G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2} \tag{14}$$

$$m_B \leq \frac{\beta[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} \tag{15}$$

$$m_A \geq \frac{[G(q_A^*) + H(e_A^*)]}{\widehat{x}^*} \tag{16}$$

$$m_B \geq \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}. \tag{17}$$

A sufficiently big value of m_A can always be chosen to satisfy both (14) and (16), so the first best is implementable only if there is an m_B satisfying both (15) and (17). This is equivalent to

$$\frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*} \leq \frac{\beta[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2},$$

which simplifies to (11) in the Proposition.■

Proof of Corollary 2: Consider Provider A. The value of m_A must satisfy both (14) and (16). The minimum value for m_A must be the larger of the right-hand side expressions of (14) and (16). We now show that the right-hand side expression in (14) is the larger one. The difference between them, after simplification, is:

$$\begin{aligned} & \frac{G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2} - \frac{[G(q_A^*) + H(e_A^*)]}{\widehat{x}^*} \\ = & \frac{[G(q_A^*) + H(e_A^*)]/2 - [G(q_B^*) + H(e_B^*)]\widehat{x}^*}{\widehat{x}^*(\widehat{x}^* - 1/2)}. \end{aligned}$$

Since the denominator is always positive ($\widehat{x}^* > 1/2$), the sign of the above is that of the numberator. That is, the right-hand side of (14) is larger than the right-hand side of (16) if and only if $\dfrac{G(q_A^*) + H(e_A^*)}{G(q_B^*) + H(e_B^*)} > 2\widehat{x}^*$.

Now by (11), it is sufficient to show that $\frac{1}{2(1-\widehat{x}^{*})} > 2\widehat{x}^{*}$. For this, we calculate

$$\begin{aligned} & \frac{1}{2(1-\widehat{x}^{*})} - 2\widehat{x}^{*} \\ = \quad & \frac{(1-2\widehat{x}^{*})^{2}}{2(1-\widehat{x}^{*})} > 0. \end{aligned}$$

Hence, the lowest possible first-best equilibrium m_A is the one in (12).

Now consider Provider B. Because m_B cannot be lower than the right-hand side of (17), the lowest possible first-best equilibrium m_B is (13).■

References

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

Milgrom, P. and Shannon, C. "Monotone Comparative Statics." *Econometrica*, Vol. 62 (1994), pp. 157-180.