Tiered and Value-based Health Care Networks

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

January 2014

Abstract

This paper studies the design of health care networks. Information systems for consumers, payment mechanisms for providers, and copayments for consumers are simultaneously considered. The optimal design aims to implement efficient provider qualities and cost-reduction efforts, and efficient allocation of consumers across different providers. For information system, we consider Tiered and Valued-based health care networks. In a Tiered Network, a provider is assigned a quality-cost designation, say, Excellent, if its quality and cost effort are above some thresholds; otherwise, it is assigned a Standard designation. In a Valued-based Network, a quality-cost index is constructed for a provider. For payment mechanisms, we consider cost reimbursement and prospective payments. Consumer copayments may be based on a provider's quality and cost. We show that the first best can be implemented in Tiered and Value-based Networks under either cost reimbursement or prospective payment. We compare the implementation costs across different designs.

Keywords: tiered network, value-based network, health plan, prospective payment, cost reimbursement

Acknowledgement: For their comments and suggestions, we thank seminar participants at Boston University, European University Institute, IUPUI, and the 14th European Health Economic Workshop in Lyon. The second author received financial and research support from the Max Weber Programme at the European University Institute.

1 Introduction

Health care markets suffer from a number of market failures. First, there is moral hazard on the demand side. Due to health insurance, consumers pay subsidized prices, and tend to overvalue service benefit and undervalue cost efficiency. Second, and arguably more serious, providers' service quality and cost efficiency are difficult to contract upon, and may be unobserved by consumers. Information design in health care networks is the latest innovation to address these market failures. In this paper we present a model of information-based networks, and study how they can complement supply-side and demand-side financial incentives to overcome market failures.

We consider two information systems. In a *Tiered Network*, a health plan uses providers' quality and cost information to sort providers into different tiers, which serve as quality-cost indicators. For example, after a provider's quality and cost information has been assessed, it is designated as either *Standard* or *Excellent*. In a *Value-based Network*, a health plan uses providers' quality and cost information to construct value indexes, which inform consumers about qualities. For example, a provider's value index of 80 may be the simple average of its quality score of 90 and efficiency score of 70. In both networks, a health plan can also use quality and cost information to set consumer cost shares. These networks are now common.[1] For example, in 2011, Blue Cross Blue Shield of Massachusetts introduced *Hospital Choice Cost Sharing* networks and Aetna began offering *Choose and Save* networks in 23 states.[2] In 2013, 23% of employers in the United States included tiered networks in their health plans (Kaiser, 2013).

For a Tiered Network or a Valued-based Network, we study how the two most common payment mechanisms, cost reimbursement and prospective payment, can be used for the implementation of providers' service quality and cost-reduction efforts. We also let the competing providers be heterogenous; some are more technologically efficient than others. This asymmetry is a necessary element in a model for studying mechanisms that rank providers.

We derive some principles of good network design. First, tiers and value indexes are effective for aligning

[1] See, for example, Robinson (2003).

[2] http://www.bluecrossma.com/plan-education/medical/hccs/.

providers' quality and cost incentives, precisely because they bundle quality and cost information. Indeed, we construct optimal tiers and value indexes. Second, consumers need not bear very high costs for optimal service utilization. In fact, we derive the lowest cost shares for efficiency, which generally depend on the quality and cost differences among providers. Finally, our results on network design change the conventional wisdom about the difference in incentive properties between cost reimbursement and prospective payment. We actually show that either cost reimbursement or prospective payment can implement efficient quality and cost efforts in a network.

In our model, two providers produce horizontally and vertically differentiated health services for consumers covered by a health plan. A provider's unit cost increases in service quality, but can be reduced by an investment. Both quality and cost-reduction investments are costly, but providers differ in investment efficiency. One provider's investment cost is a fraction of the other's. Quality and cost-reduction efforts are not contractible.

Each of a set of covered consumers picks one of the two providers for service. However, consumers cannot directly observe providers' qualities (or costs). They have to obtain such information from the health plan. We postulate that the health plan can obtain information about providers' qualities and cost efforts, and that it chooses how to report the information to consumers. Although the health plan could have disclosed all available information to consumers, we show that partial disclosure of quality and cost information can motivate providers to invest in quality and cost efforts.[3]

We identify disclosure policy with network design. In a Tiered Network, the health plan assigns a provider to a tier if its quality and cost effort are above certain thresholds. The network therefore chooses thresholds and discloses whether a provider's quality and cost effort are above those thresholds. The health plan also sets consumers' copayment for using a provider according to that provider's tier. In a Value-based Network, the health plan constructs a weighted average of a provider's quality and cost effort, and reports that index to consumers. The network therefore chooses weights on quality and cost and discloses summary statistics of a provider's services. The health plan can also set consumers' copayment for using a provider according

[3] The usual argument for partial disclosure is that limited expertise and cognitive capacities make full disclosure impractical. Although we are sympathetic to this view, we have chosen to ignore "bounded rationality" in this paper.

to how much the provider costs the plan.

In a Tiered Network, a strategic provider will attain any tier by choosing quality and cost efforts just satisfying the thresholds. Changing the definition of a tier changes the provider's quality and cost-effort choices. In a Value-based Network, a strategic provider will attain a certain level of the index to maximize profit, taking into account consumers' inference about quality from the index. The health plan must anticipate this provider reaction when constructing the index.

The health plan also sets copayments to induce efficient service utilization. What matters is consumers' incremental cost between providers. The efficient consumer incremental cost should let them internalize providers' cost difference. In a Tiered Network, the health plan can set the Excellent-tier copayment to zero, and the Standard-tier copayment to the unit-cost difference between providers in the two tiers. The same principle applies to a Value-based Network.

The health plan may use either cost reimbursement or perspective payment to implement quality and cost targets in Tiered and Value-based Networks. This is because both tiers and indexes can create complementary incentives for quality and cost efforts. However, implementation requires the health plan to carefully coordinate its provider payment and information systems. Moreover, because quality and cost incentives are different under prospective and cost-reimbursement systems, the costs for implementing the same quality and cost targets differ. We identify service cost structures which make one payment system less expensive than the other.

Our network designs are information based. In 2010, about 44% of consumers have access to health-plan sponsored hospital quality reports (Christianson et al., 2010). Dranove and Sfekas (2008) find that quality reports do affect consumer choices after consumer prior information is accounted for. Recently, Medicare and other sponsors also disclose providers' cost-efficiency information. However, researchers are skeptical about its effect on consumer choice (Mehrotra et al. 2012). Our model shows that demand response to cost disclosure critically depends on how quality and cost information are conveyed.

We propose that quality and cost information be bundled into value measures. The Massachusetts Group Insurance Commission is one of the first purchasers to adopt this approach. The Commission started offering tiered networks to more than 267,000 state employees and retirees in 2004 (Alteras and Silow-Carroll, 2007). In these tiered networks, health plans use quality and cost-efficiency information to assign providers to one of Tiers 1, 2, and 3. For example, a primary care physician is assigned to Tier 1 only if its quality score is above B on a scale from A to C and its efficiency score is above average.[4] Consumers are informed about a provider's tier only but not the underlying quality or efficiency score. This approach is consistent with our Tiered Network. We also study value index as an alternative disclosure strategy. More important, we illustrate how tier and index construction, consumer cost sharing, and provider payment should be coordinated. To the best of our knowledge, this is the first attempt in the literature to do so.[5]

Innovations in consumer cost sharing are increasingly prominent. A number of recent papers study value-based consumer cost sharing. Thomson et al. (2013) report that value-based cost sharing has been adopted in the United States and more than 10 European countries. Pauly and Blavin (2008) study the price theory of value-based cost sharing; their analysis is based on a price-taking service market where quality and cost efficiency are fixed. They show that when consumers are fully informed, optimal cost-based and value-based copayments are identical. Valued-based copayment is lower only when consumers are imperfectly informed and underestimate service benefit. They argue that information disclosure and value-based cost sharing are substitutes. By contrast, we study copayment policy in an oligopolistic service market where both quality and cost are endogenous. In our Tiered and Value-based Networks, information disclosure and copayments together guide consumers to make efficient provider choices. We show that the optimal copayment for efficient utilization of high-value services can be as low as zero.

In our model consumers are rational. Baicker et al. (2012) study demand for health services when consumers suffer from various behavioral biases. They also find that the optimal copayment for high-value services can be zero. Our results complement theirs. Moreover, we show that zero copayment is consistent with production efficiency.

The utilization and production efficiency results distinguish our work from other papers that study

[4] http://www.unicarestateplan.com/pdf/CPIPhysicianBrochureFY13.pdf

[5] The three policy instruments have been studied extensively but separately. For recent surveys of the three strands of literature, see Dranove and Jin (2010), Léger (2008), and McGuire (2011).

oligopolistic provider competition. Beitia (2003) and Levaggi (2005) also consider a health plan contracting with two providers that produce horizontally differentiated services. In their papers quality and cost efficiency are not contractible either. However, they let consumers always observe quality and pay zero copayment. In both papers, first-best utilization is unattainable because the health plan lacks the copayment instrument to direct consumer choices. In their second-best solutions, qualities are distorted from the social optimal to minimize consumption inefficiency. By contrast, we obtain conditions for efficiency; this illustrates that information disclosure in network design, provider payment, and consumer copayment should be considered simultaneously.[6]

The paper is organized as follows. Section 2 sets up the model of a health plan incentivizing providers to service enrollees. We study Tiered and Valued-based Networks, respectively, in Sections 3 and 4. There, for each information-based network, we show the implementation of first-best qualities and cost efforts under cost reimbursement and prospective payment, and compare the health plan's implementation costs under the two payment systems. In Section 5, we then compare the implementation costs of Tiered and Value-based Networks. Finally, Section 6 concludes. All proofs are in the Appendix.

2 The model

We now set up a model of health care networks. 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 $C(q, e)$. The unit cost function is strictly increasing in quality and strictly

[6] Ma (1994) shows that prospective payment can achieve first-best quality and cost efficiency. Ma and Mak (2013) show that cost reimbursemnt together with information disclosure can implement the same outcome. Both papers consider a purchaser contracting with a single provider.

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 effort. If Provider A takes quality and cost effort (q_A, e_A), its fixed cost is $G(q_A) + H(e_A)$, where G and H are both strictly increasing and strictly convex functions. If Provider B takes quality and cost effort (q_B, e_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. The inefficiency may be due to higher management costs, or tighter local market conditions which drive up effort and quality costs, etc. 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

The total mass of consumers covered by the health plan is normalized at 1. Each consumer is described by a valuation 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. We use the common Hotelling, horizontal product-differentiation structure to describe consumers' proclivity towards the two providers. Consumers are uniformly distributed on the $[0, 1]$ interval. A nonnegative parameter τ measures 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 copayment to use a provider. We will expand on copayment 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, so the market is covered.

We assume that consumers have the same valuation on quality. Our interpretation is that we consider the subset of consumers within a health plan 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, 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 less the production cost. Let the health plan assign consumers with $x \leq \widehat{x}$ to Provider A, and the remaining consumers to Provider B. Social welfare is

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

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. (All proofs are in the Appendix.)

Lemma 1 *In the first best, Provider A sets a higher quality, a higher cost effort, and serves more consumers than Provider B. That is, if q_A^*, q_B^*, e_A^*, e_B^*, and $\widehat{x}^*$ denote the first-best qualities and efforts by Providers A and B and the consumer allocation across them, then $q_A^* > q_B^*$, $e_A^* > e_B^*$, and $\widehat{x}^* > 1/2$. Furthermore,* *they satisfy the following:*

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

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

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

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

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

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, high effort, 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 services 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 the two most common forms of provider payment. Under *prospective payment*, the health plan pays p_i to Provider i for a unit of service. Under *cost reimbursement*, the health plan pays any *ex post* unit cost $C(q_i, e_i)$ plus a (positive) margin m_i to Provider i for a unit of service. In this paper, we only consider these two forms of payments in the analytical models.[7]

[7] Consumer cost heterogenity can be readily incorporated into this framework. See, for example, Ma (1994), Ma and Mak (2013).

Consumers' choices of providers are also noncontractible. As mentioned above, the health plan may impose copayments. Consumers do not observe 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 in order 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 Section 3.) 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 a value index equal to a linearly weighted sum of the quality and the cost effort. (The value index will be formally defined in Section 4.) The health plan then discloses providers' quality indexes to consumers. The health plan also sets consumers' copayments for obtaining services from 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

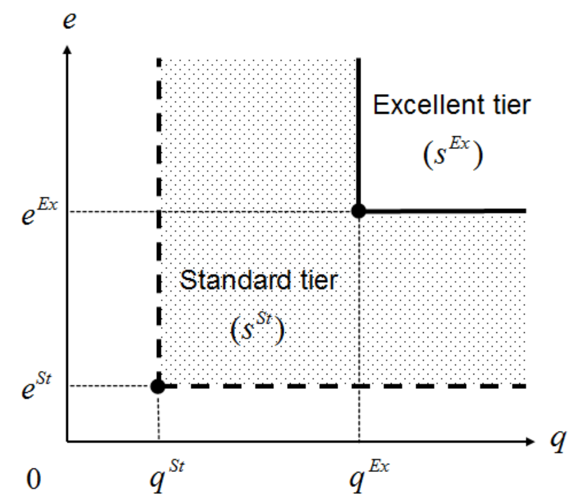

Figure 1: Tiers and copayments

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-shaped region in Figure 1. A consumer pays $s^{Ex} \geq 0$ 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 commits to reimbursing Provider i's operating cost, and sets the margin m_i, $i = A, B$. 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 qualities and cost-reduction efforts simultaneously.

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

Stage 4: Consumers learn the providers' designated tiers, and decide on obtaining services from one of them.

Consumers do not observe the providers' qualities and efforts. The health plan does observe these, and assign providers to different tiers. In Stage 4, consumers form beliefs about qualities and efforts based on providers' tiers. 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 designations, consumers believe that q'_A and q'_B are service qualities of Providers A and B, respectively. Consumers believe that utilities from Providers A and B are, respectively, $vq'_A - s_A - \tau x$ 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 large 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 that 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 in a Tiered Network. 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 (and its proof is omitted).

Corollary 1 *Consider cost reimbursement in a Tiered network. 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 Stage 2, Providers A and B choose equilibrium (q_A^*, e_A^*) and (q_B^*, e_B^*), respectively. In Stage 3 the health plan then reports that Provider A belongs to the Excellent tier and Provider B to the Standard tier. By Lemma 2, consumers' demand for services from 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 $\widehat{x}^*$ 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 the less efficient Provider B.

Lemma 3 *Suppose the first best is implementable as a perfect-Bayesian equilibrium in a Tiered Network. The lowest, nonnegative 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 of $C(q_A^*, e_A^*)$ and $C(q_B^*, e_B^*)$. Lemma 1 prescribes 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 parameters 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

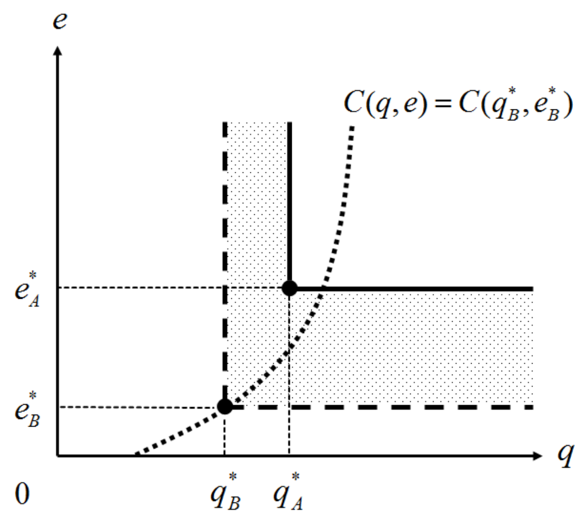

Figure 2: Equilibrium tiers and copayments

at the first-best unit cost level.[8] 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 at the Standard tier.

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 Corollary 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. Therefore, their market shares are $\widehat{x}^*$ and $1 - \widehat{x}^*$, respectively. This shows up as the first row and the 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-column cells in Table 1.

[8] 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		
		(q_A^*, e_A^*)	(q_B^*, e_B^*)	$(0,0)$
Provider A	(q_A^*, e_A^*)	$1/2, 1/2$	$\widehat{x}^*, 1-\widehat{x}^*$	$1, 0$
	(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

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 a unilateral deviation profitable.

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

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

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

and nonnegative profit constraints:

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

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

In a first-best equilibrium, Provider A chooses (q_A^*, e_A^*). The incentive constraint (7) says that Provider A should not profit by deviating to (q_B^*, e_B^*) and capturing $1/2$ of the market (see Table 1). The incentive constraint (8) has the corresponding meaning for Provider B. 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? In other words, which of the incentive and nonnegative profit constraints are more binding?

By Lemma 1, $\widehat{x}^* > 1/2$ and $G(q_A^*) + H(e_A^*) > G(q_B^*) + H(e_B^*)$. Therefore, a sufficiently big m_A can satisfy both (7) and (9), so Provider A has no incentive to deviate from the Excellent tier if m_A is sufficiently big. The incentive and nonnegative profit constraints for Provider B work differently. Conditions (8) and (10) constrain m_B in opposite directions. The incentive constraint (8) requires m_B to be small, so that deviating to (q_A^*, e_A^*) and getting the additional $(\widehat{x}^* - 1/2)$ consumers is unprofitable. The nonnegative profit constraint (10) requires m_B to be big so that Provider B remains active. Combining both constraints, we obtain

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

which yields the condition in the next Proposition.

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

$$G(q_A^*) + H(e_A^*) \geq \frac{G(q_B^*) + H(e_B^*)}{2(1 - \widehat{x}^*)}. \tag{12}$$

Inequality (12) in Proposition 1 says that in the first best, the total fixed cost of quality and cost effort for the more efficient Provider A must still be quite larger than the less efficient Provider B. This is a direct consequence of the requirement that m_B stays within the bounds in (11) for the first best.

We next turn to the cost of first-best implementation. The health plan's total payment to the providers in a first-best equilibrium is $[m_A + C(q_A^*, e_A^*)]\widehat{x}^* + [m_B + C(q_B^*, e_B^*)](1 - \widehat{x}^*)$. Therefore, the health plan minimizes payment to the providers when the margins are chosen to be the smallest required for implementation. The value of m_B must satisfy (11), so the minimum feasible value must be $\frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}$. The value of m_A must satisfy (7) and (9). Our next result shows that the incentive constraint (7) is the more binding one.

Corollary 2 *Consider the implementation of the first best in a Tiered Network. The cost-minimizing margins are*

$$m_A = \frac{[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} \tag{13}$$

$$m_B = \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}. \tag{14}$$

The minimum margin m_A in (13) for implementing the first best is obtained by a binding Provider A incentive constraint (7). Provider A's nonnegative profit constraint (9) is slack. The more efficient Provider

A must make a profit in a first-best equilibrium. This is a consequence of condition (12) in Proposition 1; this requires $\widehat{x}^*$ to be not much larger than 1/2, and $G(q_A^*) + H(e_A^*)$ larger than $G(q_B^*) + H(e_B^*)$. Condition (12) ensures that the more efficient Provider A makes a positive profit if it deviates to (q_B^*, e_B^*). Conversely, the value of m_B in (14) comes from a binding Provider B nonnegative profit constraint (10). Condition (12) in Proposition 1 says that at that value the incentive constraint (8) is slack: the less efficient Provider B makes a negative profit if it deviates to (q_A^*, e_A^*). Therefore, Provider B's profit is zero in the first-best equilibrium.

3.3 Prospective payment and first best

We now study the implementation of the first best under prospective payment. Here, Providers A and B receive p_A and p_B, respectively, for each unit of service. Each provider fully internalizes the service unit costs and fixed costs. Because of this internalization, the specification of tiers according to (q^{Ex}, e^{Ex}) or (q^{St}, e^{St}) in Lemma 2 can be relaxed. We can alternatively define a tier based on a provider's quality only. For example, a provider belongs to the Excellent tier if its quality is at least q^{Ex}; a provider belongs to the Standard tier if its quality is at least q^{St}. However, for ease of exposition, we will continue to use the same tier construction as in the last subsection, as in Corollary 1.

Given the same tier construction, an active provider's choice of quality and effort continue to be either (q_A^*, e_A^*) or (q_B^*, e_B^*) in a perfect-Bayesian equilibrium. We specify consumers' beliefs in the same way. That is, in a perfect-Bayesian equilibrium, consumers believe that a provider's quality and cost effort are the minimum that would qualify for its tier designation. We also let the health plan use the same copayments for first-best implementation as in Lemma 3 in the last subsection.

The first best is implementable if there are prospective prices p_A and p_B that satisfy the incentive constraints:

$$[p_A - C(q_A^*, e_A^*)]\widehat{x}^* - [G(q_A^*) + H(e_A^*)] \quad \geq \quad [p_A - C(q_B^*, e_B^*)]/2 - [G(q_B^*) + H(e_B^*)] \qquad (15)$$

$$[p_B - C(q_B^*, e_B^*)](1 - \widehat{x}^*) - \beta[G(q_B^*) + H(e_B^*)] \quad \geq \quad [p_B - C(q_A^*, e_A^*)]/2 - \beta[G(q_A^*) + H(e_A^*)] \qquad (16)$$

and nonnegative profit constraints:

$$[p_A - C(q_A^*, e_A^*)]\widehat{x}^* - [G(q_A^*) + H(e_A^*)] \geq 0 \tag{17}$$

$$[p_B - C(q_B^*, e_B^*)](1 - \widehat{x}^*) - \beta[G(q_B^*) + H(e_B^*)] \geq 0. \tag{18}$$

Incentive constraints (15) and (16) say that Providers A and B, respectively, prefer to achieve the Excellent and Standard tiers. The nonnegative profit constraints (17) and (18) ensure that both providers are active.

Because $\widehat{x}^* > 1/2$, both (15) and (17) can be satisfied when p_A is sufficiently high. A high prospective price p_A encourages the more efficient Provider A to attain the Excellent Tier to get a larger market share. By contrast, again because $1 - \widehat{x}^* < 1/2$, Provider B's incentive constraint (16) requires that p_B be small, while the nonnegative profit constraint (18) requires p_B to be large. By combining (16) and (18), we get the condition in the following Proposition.

Proposition 2 *Consider prospective payment in a Tiered Network. Let the tier and copayment policies be given by those in Corollary 1 and Lemma 3. The first best is implemented as a perfect-Bayesian equilibrium if and only if*

$$G(q_A^*) + H(e_A^*) \geq \frac{G(q_B^*) + H(e_B^*)}{2(1 - \widehat{x}^*)} + \frac{C(q_B^*, e_B^*) - C(q_A^*, e_A^*)}{2\beta}. \tag{19}$$

Under prospective payment, each provider internalizes its variable cost $C(q, e)$. This accounts for the extra term in (19) in Proposition 2 compared to (12) in Proposition 1. Provider A is more efficient, and in the first best puts in more quality and cost effort, so the comparison between variable costs of the two providers is ambiguous in general. Indeed, (19) implies (12) if and only if $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$. When Provider A's first-best variable cost is lower than Provider B's, $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$, the first best is implementable by prospective payment only if it is implementable by cost reimbursement. Symmetrically, (12) implies (19) if and only if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$, so the first best is implementable by prospective payment only if it is implementable by cost reimbursement when Provider A's variable cost is higher.

The health plan pays $p_A\widehat{x}^* + p_B(1 - \widehat{x}^*)$ to implement the first best. The minimum feasible value of p_B must satisfy the binding nonnegative profit constraint (18). The minimum feasible value of p_A must satisfy both the incentive constraint (15) and the nonnegative profit constraint (17). However, either one of these may be the binding constraint. Indeed, if Provider A earns a strictly positive profit by deviating to the Standard tier (so that the right-hand side of (15) is strictly positive), then the nonnegative constraint (17) is slack. The converse is also true.

Whether Provider A's incentive constraint or nonnegative profit constraint binds depends on the comparison of the right-hand side of (15) and the left-hand side of (17). Clearly, the incentive constraint (15) implies the nonnegative profit constraint (17) if and only if

$$C(q_A^*, e_A^*) + \frac{G(q_A^*) + H(e_A^*)}{\widehat{x}^*} \geq C(q_B^*, e_B^*) + \frac{G(q_B^*) + H(e_B^*)}{1/2}. \tag{20}$$

Given consumers' equilibrium responses to tier designation, the left-hand side of (20) is Provider A's average service cost when quality and cost effort qualify for the Excellent tier, while the right-hand side is the average cost when Provider A deviates to the Standard tier. Condition (20) requires $\widehat{x}^* - 1/2$ and $C(q_B^*, e_B^*) - C(q_A^*, e_A^*)$ to be small, and $G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)$ to be large.

For Provider B, condition (19) in Proposition 2 says that its incentive constraint (16) is slack. The binding nonnegative profit constraint (18) determines the minimum feasible value of p_B. The next Corollary shows how inequality (20) determines the minimum costs of first-best implementation.

Corollary 3 *Consider the implementation of the first best in a Tiered Network. Suppose that the incentive constraint (15) implies the nonnegative profit constraint (17) (equivalent to (20) satisfied), the cost-minimizing prospective price for Provider A is*

$$p_A = \frac{C(q_A^*, e_A^*)\widehat{x}^* - C(q_B^*, e_B^*)/2 + G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2}. \tag{21}$$

Suppose that the nonnegative profit constraint (17) implies the incentive constraint (15) (equivalent to (20) violated), the cost-minimizing prospective price for Provider A is

$$p_A = C(q_A^*, e_A^*) + \frac{[G(q_A^*) + H(e_A^*)]}{\widehat{x}^*}. \tag{22}$$

The cost-minimizing prospective price for Provider B is

$$p_B = C(q_B^*, e_B^*) + \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}. \tag{23}$$

3.4 Payment mechanism and implementation cost

We now compare the implementation costs under cost reimbursement and prospective payment. Clearly, the comparison is meaningful only when the first best can be implemented by both mechanisms. According to Propositions 1 and 2, this is true when conditions (12) and (19) are satisfied. The resource costs under the first best are predetermined, so the implementation cost comparison has to do with providers' profits. In other words, we should compare the margin-unit-cost combinations under cost reimbursement and the prospective prices. By Corollaries 2 and 3, the less efficient Provider B always makes a zero profit, so $m_B + C(q_B^*, e_B^*) = p_B$ (see also (14) and (23) in the two Corollaries). The comparison therefore boils down to the values of $m_A + C(q_A^*, e_A^*)$ and p_A in the two Corollaries.

By Corollary 2, the minimum margin for Provider A is in (13). By Corollary 3, the minimum prospective price for Provider A is either (21) or (22), depending on whether the incentive constraint binds or the nonnegative profit constraint binds. First, suppose that Provider A's price is given by (21). Combining this and (13), we obtain (details in the proof)

$$p_A - [m_A + C(q_A^*, e_A^*)] = \frac{C(q_A^*, e_A^*) - C(q_B^*, e_B^*)}{2\widehat{x}^* - 1}. \tag{24}$$

The equality says that $p_A \geq [m_A + C(q_A^*, e_A^*)]$ if and only if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$.

Under prospective payment, when deviating from (q_A^*, e_A^*) to (q_B^*, e_B^*), Provider A internalizes the change in unit cost. This internalization is absent under cost reimbursement. Therefore, when $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$, a deviation from (q_A^*, e_A^*) to (q_B^*, e_B^*) is more profitable under prospective payment. The health plan has to pay a higher price to keep Provider A from deviating to (q_B^*, e_B^*). Symmetrically, if $C(q_A^*, e_A^*) < C(q_B^*, e_B^*)$, the same deviation to (q_B^*, e_B^*) is less profitable under prospective payment. The health plan only has to pay a lower price.

Second, suppose that Provider A's price is given by (22). Here, Provider A makes a zero profit under prospective payment. However, according to Corollary 2, Provider A always makes a strictly positive profit under cost reimbursement. (The proof of that Corollary establishes that the minimum margin is obtained from a binding incentive constraint while the nonnegative profit constraint remains slack.) Therefore, we must have $p_A < [m_A + C(q_A^*, e_A^*)]$, so that implementation cost is lower under prospective payment.

Now, Provider A's price being given by (22) requires the violation of (20). As the proof of the next Proposition shows, this, together with the necessary condition of implementation by prospective payment implies $C(q_B^*, e_B^*) > C(q_A^*, e_A^*)$.

Proposition 3 *Suppose the first best can be implemented by cost reimbursement and prospective payment in a Tiered Network. The health plan minimizes the payment to providers by using cost reimbursement if and only if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$.*

In this section we have compared cost reimbursement and prospective payment in Tiered Networks. We have focused on the implementation of the first best. In fact, our results can be applied to the implementation of other qualities, cost efforts, and assignments of consumers across the providers. In the first best, Provider A chooses higher quality and cost effort, and serves more consumers than Provider B. These are the only properties used for all the lemmas, propositions, and corollaries here. If we replace q_A^*, e_A^*, q_B^*, e_B^*, and $\widehat{x}^*$ by $\widetilde{q}_A$, $\widetilde{e}_A$, $\widetilde{q}_B$, $\widetilde{e}_B$, and $\widetilde{x}$ in these lemmas, propositions, and corollaries, the statements remain correct when $\widetilde{q}_A > \widetilde{q}_B$, $\widetilde{e}_A > \widetilde{e}_B$, and $\widetilde{x} > 1/2$. This can be verified by inspection.

Cost and quality efficiencies differ among the providers, and this yields an asymmetric first best. Suppose that, for whatever reason, a health plan wants to implement an allocation different from the first best. Our results continue to hold as long as the allocation exhibits the same kind of asymmetry, namely, the more efficient provider puts in more quality and cost effort, and serves more consumers. It is hard for us to imagine situations in which any health plan would want to implement allocations that favor the less efficient provider. Therefore, we do not consider implementation of these uninteresting allocations.

4 Value-based Network

4.1 Network structure and extensive form

In this section we consider the implementation of the first best in Value-based Networks. Here, the health plan constructs a value index for each provider based on its quality and cost-reduction effort, and continues to pay the providers by either cost reimbursement or prospective payment. A value index is a weighted average of a provider's quality and cost effort. Suppose that Provider i, $i = A, B$, chooses quality q_i and cost effort e_i. As in the previous section, these choices are noncontractible, but the health plan can observe them, and disclose some information about them. The value index for Provider i, $i = A, B$, is $I_i(\theta_i) \equiv \theta_i q_i + (1 - \theta_i)e_i$, where $0 \leq \theta_i \leq 1$, is the weight on quality. The health plan can commit to the weights θ_A and θ_B, as well as payment parameters and consumer copayments before the providers make decisions.

The extensive form for the Value-based Network game is as follows.

Stage 1: The health plan sets (θ_A, θ_B) and copayments (s_A, s_B). 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 qualities and cost-reduction efforts simultaneously.

Stage 3: The health plan observes the providers' chosen qualities and efforts, and reports the value index $I_i(\theta_i) = \theta_i q_i + (1 - \theta_i)e_i$ of each Provider $i = A, B$.

Stage 4: Consumers learn the providers' value indexes (but not providers' qualities and cost efforts), and decide between obtaining services from Provider A and Provider B.

We study perfect-Bayesian equilibria under cost reimbursement and prospective payment in the next two subsections.

4.2 Cost reimbursement and first best

In Tiered Networks, a health plan would only announce if a provider qualifies for one of two possible tiers. In Value-based Networks, a provider's value index can take on any nonnegative number. Each level of index $I_i(\theta_i)$ can be achieved by many combinations of q_i and e_i. Given such indexes, what do consumers believe about providers' qualities and cost efforts in a perfect-Bayesian equilibrium?

Suppose that equilibrium value indexes of Providers A and B are, respectively, $I_A(\theta_A)$ and $I_B(\theta_B)$, and that in equilibrium, Provider A serves $\widehat{x}$ of consumers while Provider B serves the rest. The next Lemma shows that in any perfect-Bayesian equilibrium, Provider i must choose the profit-maximizing q_i and e_i to achieve index $I_i(\theta_i)$.

Lemma 4 *Consider cost reimbursement in a Valued-based Network. Suppose that in equilibrium Provider A serves $\widehat{x}$ of consumers while Provider B serves the rest. Equilibrium qualities and cost efforts (q_A, e_A) and (q_B, e_B) must solve*

$$\max_{q'_A, e'_A} m_A \widehat{x} - [G(q'_A) + H(e'_A)]$$
$$\textit{subject to} \quad \theta_A q'_A + (1-\theta_A)e'_A = \theta_A q_A + (1-\theta_A)e_A$$

and

$$\max_{q'_B, e'_B} m_B(1-\widehat{x}) - [G(q'_B) + H(e'_B)]$$
$$\textit{subject to} \quad \theta_B q'_B + (1-\theta_B)e'_B = \theta_B q_B + (1-\theta_B)e_B.$$

Hence, (q_i, e_i), $i = A, B$, satisfy

$$\frac{G'(q_i)}{H'(e_i)} = \frac{\theta_i}{1-\theta_i}. \tag{25}$$

As in Tiered Networks, a provider will attempt to achieve a level of the value index in the most profitable way. In equilibrium, consumers must believe that profit-maximizing qualities and efforts are used to achieve an index. Once providers' index levels are given, demands, and hence revenues, are fixed. Then Lemma 4 implies that in equilibrium q_i and e_i must minimize the total fixed cost to achieve index $I_i(\theta_i)$. Condition (25) is the solution to this minimization problem. It says that the ratio of marginal fixed cost $G'(q_i)/H'(e_i)$ must be set equal to the ratio of quality and cost-effort weights $\theta_i/(1-\theta_i)$. In other words, the marginal contributions of quality and cost effort to achieve an index must be in-line to their respective marginal costs.

Figure 3 illustrates Provider i's cost-minimization problem. The downward sloping straight line contains (q, e) pairs that have the same level of value index I_i with the weight on quality being set at θ_i. The iso-cost curve plots (q, e) pairs that have the same total fixed cost; it is concave to the origin because G and H are strictly convex. The tangency point $(q_i(\theta_i), e_i(\theta_i))$ shows the cost-minimizing quality and cost effort at $I_i(\theta_i)$. For a fixed θ_i, the dotted "expansion path" from the origin describes the cost-minimizing qualities

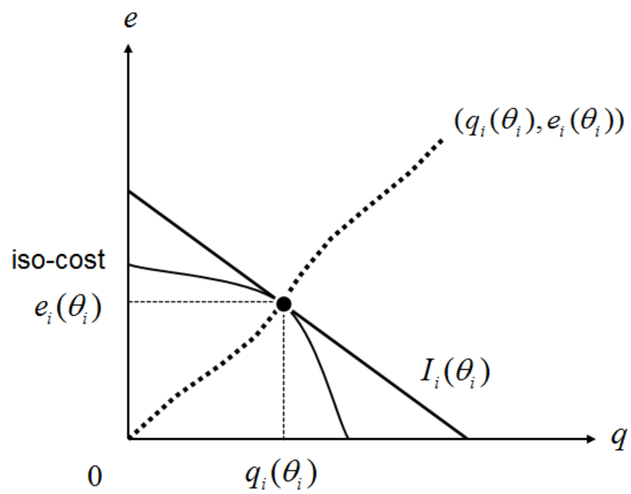

Figure 3: Cost-minimizing quality and cost effort

and cost efforts that achieve different levels of value index $I_i(\theta_i)$ in a cost-minimizing way. When the value of θ_i changes, the entire expansion path will pivot around the origin. For example, if θ_i increases so that more weight is given to quality, then more quality will be chosen to achieve any given level of the index, so the dotted expansion path will rotate in a clockwise direction.

To implement the first best, the health plan has to choose index weights θ_A and θ_B so that the first best is on the providers' expansion paths, or equivalently (25) holds at the first-best qualities and efforts.

Corollary 4 *Consider cost reimbursement in a Value-based Network. The first best is implementable as a perfect-Bayesian equilibrium only if the health plan sets index weights* $\theta_A = \dfrac{G'(q_A^*)}{G'(q_A^*) + H'(e_A^*)}$ *and* $\theta_B = \dfrac{G'(q_B^*)}{G'(q_B^*) + H'(e_B^*)}$.

The health plan also sets copayments s_A and s_B in Stage 1. Suppose that in Stage 2, Providers A and B choose (q_A^*, e_A^*) and (q_B^*, e_B^*), respectively. By Lemma 4, consumers can correctly infer qualities from the value indexes. Their demand for services from Provider A is $1/2 + (1/2\tau)[v(q_A^* - q_B^*) - (s_A - s_B)]$. To implement the first-best consumer allocation, the copayment differential $s_A - s_B$ has to internalize the difference in unit costs $C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$. Therefore, the lowest, nonnegative s_A and s_B are, respectively, identical to s^{Ex} and s^{St} in the optimal Tiered Network (see Lemma 3). Consumers' copayments remain the same whether the first best is implemented under Tiered or Value-based Networks; it follows that they obtain exactly the same utilities in each network.

Next, we derive the margins m_A and m_B for first-best implementation. By Lemma 4, each provider has to choose among combinations of qualities and cost efforts under constraint (25), which describe what can arise in equilibrium. Given m_A and m_B, the equilibrium in Stage 2 is given by the solutions of the two following programs:

$$\max_{q_A,e_A} \; m_A\left[\frac{1}{2}+\frac{v(q_A-q_B)-(s_A-s_B)}{2\tau}\right]-[G(q_A)+H(e_A)] \tag{26}$$
$$\text{subject to} \;\; \frac{G'(q_A)}{H'(e_A)}=\frac{\theta_A}{1-\theta_A}$$

and

$$\max_{q_B,e_B} \; m_B\left[\frac{1}{2}+\frac{v(q_B-q_A)-(s_B-s_A)}{2\tau}\right]-\beta[G(q_B)+H(e_B)] \tag{27}$$
$$\text{subject to} \;\; \frac{G'(q_B)}{H'(e_B)}=\frac{\theta_B}{1-\theta_B}.$$

In Stage 1, the health plan must choose m_A and m_B so that the first best is a continuation equilibrium.

Proposition 4 *Consider cost reimbursement in a Valued-based Network. Let the index weights be given by Corollary 4 and let the copayments satisfy $s_A - s_B = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$. The first best is implemented as a perfect-Bayesian equilibrium if and only if*

$$m_A = \frac{1}{v/2\tau}\left[G'(q_A^*)+H'(e_A^*)\times\frac{d\ln G'(q_A^*)}{d\ln H'(e_A^*)}\right] \tag{28}$$
$$m_B = \frac{\beta}{v/2\tau}\left[G'(q_B^*)+H'(e_B^*)\times\frac{d\ln G'(q_B^*)}{d\ln H'(e_B^*)}\right]. \tag{29}$$

Proposition 4 presents a unique pair of margins (m_A, m_B) for first-best implementation. To obtain these margins, we begin with the constraints in (26) and (27). Each implicitly defines the equilibrium e_i, $i = A, B$, as a function of q_i, say $e_i(q_i)$ (and this is illustrated by the dotted expansion path in Figure 3). The slope of this function, $\frac{1-\theta_i}{\theta_i}\frac{G''(q_i)}{H''(e_i)}$ (derived in the Appendix), describes how, in equilibrium, e_i changes with respect to q_i. Consider Provider A. To induce the first-best quality and effort, the health plan has to set the margin m_A such that the provider's marginal benefit of quality investment, $m_A \times v/2\tau$, is equal to the marginal fixed cost due to investment, $G'(q_A)+H'(e_A)\times e_A'(q_A)$, evaluated at the first best. The expressions for m_A in Proposition 4 is obtained by solving this equation. The expression for m_B is derived in the same way.

4.3 Prospective payment and first best

We now study first-best implementation under prospective payment. The provider payment literature has shown that health plans and providers have aligned cost-efficiency incentives under complete information. Our next result says that when the health plan constructs value indexes under prospective payment, it assigns zero weight to cost effort in order to implement the first best. Our result reaffirms those in the extant literature.

Lemma 5 *Consider prospective payment in a Value-based Network. The first best is implementable as a perfect-Bayesian equilibrium only if the health plan sets $\theta_A = \theta_B = 1$.*

Under prospective payment, a provider fully internalizes the benefit of cost reduction. There is no need for extra incentive for cost effort. In other words, simply ignoring cost effort in the value index, disclosing the quality, is sufficient. Under prospective payment, a value index should put all weights on quality.

In the last subsection, we have established that consumers have to pay the same minimum copayments whenever the first best is implementable. But as the next Proposition shows, the optimal prospective prices and cost margins in Value-based Networks are chosen to address distinctive incentive problems.

Proposition 5 *Consider prospective payment in a Value-based Network. Let the index weights be $\theta_A = \theta_B = 1$ and the copayments satisfy $s_A - s_B = C(q_A^*, e_A^*) - C(q_B^*, e_B^*)$. The first best is implemented as a perfect-Bayesian equilibrium if and only if*

$$\begin{aligned} p_A &= \tau + C(q_B^*, e_B^*) + v(q_A^* - q_B^*) \\ p_B &= \tau + C(q_A^*, e_A^*) + v(q_B^* - q_A^*). \end{aligned}$$

Because the providers internalize the costs and benefits of efforts, their effort incentives are aligned with the health plan's. We only need to choose prospective prices to align providers' quality incentives. Indeed, there is a basic divergence between a provider's private quality benefit and social quality benefit. For example, if Provider A raises q_A by one unit, its gross profit increases by $[p_A - C(q_A, e_A)]\dfrac{v}{2\tau}$, where $p_A - C(q_A, e_A)$ is the gross profit margin and where $\dfrac{v}{2\tau}$ is the provider's gain in market share. However, the increase in quality also benefits Provider A's current consumers so the increase in social benefit is $v\widehat{x}$. This divergence of incentives can be likened to the problem of monopolistic quality provision in Spence (1975).

To implement the first best, the health plan must choose prices so that a provider's marginal profit due to quality is equal to the marginal social benefit. The expressions for p_A and p_B in Proposition 5 are chosen to do that. Specifically, the optimal prospective price of a provider is increased when demand becomes less responsive to quality, or when the rival provider has a higher unit cost at the first best. Moreover, Provider A is rewarded for its higher average social benefit $v(q_A^* - q_B^*) > 0$, whereas the prospective price for Provider B is reduced for its lower average social benefit $v(q_B^* - q_A^*) < 0$.

4.4 Payment mechanism and implementation cost

We now study implementation costs under the two payment systems. Unlike in Tiered Networks where only Provider A can make positive profits, both providers earn positive profits in Value-based Networks. Implementation costs take the form of provider profits. Provider i's, $i = A, B$, profit margin under prospective payment is $p_i - C(q_i^*, e_i^*)$, and m_i under cost reimbursement, so it is sufficient to compare these two values. In the next proposition we compare the values of $m_i + C(q_i^*, e_i^*)$ and p_i.

Proposition 6 *For Value-based Networks, Provider i earns a higher profit under prospective payment than cost reimbursement, $p_i > m_i + C(q_i^*, e_i^*)$, if and only if*

$$-\frac{C_q(q_i^*, e_i^*)}{C_e(q_i^*, e_i^*)} > \frac{d \ln G'(q_i^*)}{d \ln H'(e_i^*)}. \tag{30}$$

The comparison between a provider's profits under cost reimbursement and prospective payment depends on the curvatures of the unit cost C and the fixed costs G and H at the first best. Figure 4 illustrates an example where inequality (30) in Proposition 6 holds. The solid curve in the diagram is the iso-unit cost curve of Provider i at the first best (see the discussion following Lemma 3). By the implicit function theorem, the slope of this curve is $-\dfrac{C_q(q_i, e_i)}{C_e(q_i, e_i)}$ at (q_i, e_i). The dashed curve in Figure 4 plots the combinations of fixed-cost minimizing (q_i, e_i) that we have introduced in Figure 3. There, we also have shown that under Proposition 4, the slope of the dashed curve is $\dfrac{1-\theta_i}{\theta_i}\dfrac{G''(q_i)}{H''(e_i)}$ at (q_i, e_i) and can be written as $\dfrac{d \ln G'(q_i^*)}{d \ln H'(e_i^*)}$ at (q_i^*, e_i^*).

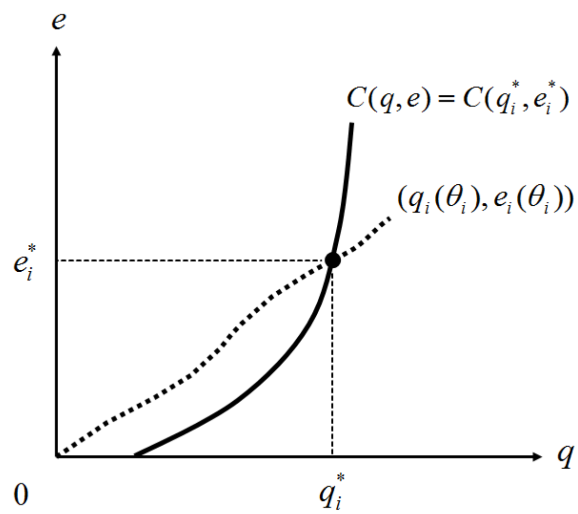

Figure 4: Comparison of providers' profits

Under prospective payment, the first-best p_i is set to make Provider i's iso-profit (net of unit cost) line tangent to the solid curve at (q_i^*, e_i^*). Similarly, under cost reimbursement, the first-best m_i is set such that Provider i's iso-profit line tangents to the dashed curve at (q_i^*, e_i^*). In Figure 4 the inequality (30) is satisfied, so the solid curve is steeper than the dashed curve at (q_i^*, e_i^*). In other words, the first-best cost effort is more costly to Provider i under prospective payment. In order to implement the first best, the health plan has to offer a higher price under prospective payment to make Provider i's iso-profit curve steeper. This means that p_i is higher than $m_i + C(q_i^*, e_i^*)$. Because the comparison between p_i and $m_i + C(q_i^*, e_i^*)$ depends on the curvatures of the cost functions at (q_i^*, e_i^*), it is possible that Provider A earns a higher profit under cost reimbursement, but Provider B earns a higher profit under prospective payment, and vice versa.

5 Optimal network

In the previous two sections, we have studied how the first best can be implemented in Tiered and Value-based Networks. By Propositions 1 and 2, the first best can be implemented in Tiered Networks only if either condition (12) or (19) is satisfied. On the other hand, by Propositions 4 and 5, the first best can always be implemented in Value-based Networks. This difference is due to the interaction between network design and providers' nonnegative profit constraints. In Tiered Networks, an active provider can only choose between

(q_A^*, e_A^*) and (q_B^*, e_B^*) (see Lemma 2 and Corollary 4). Conditions (12) and (19) are necessary conditions for the less efficient Provider B to make a nonnegative profit at (q_B^*, e_B^*). In Value-based Networks, an active provider can choose many pairs of qualities and cost efforts. Provider payments are chosen to satisfy the providers' incentive constraints, while providers' nonnegative profit constraints never bind.

When the first best can be implemented in both Tiered and Value-based Networks, Provider B's nonnegative profit constraint binds in Tiered Networks only. However, the binding constraint does not imply that the total implementation cost must be lower in Tiered Networks. To compare total implementation cost, we have to consider the total profits of Providers A and B under different networks *and* payment mechanisms. We now make this comparison assuming conditions (12) and (19) are satisfied.

Proposition 7 *Suppose that the first best can be implemented in both Tiered and Value-based Networks. If condition (30) for Provider A is violated, the health plan pays a lower cost to providers by using Tiered Networks. On the other hand, if condition (30) for Provider A holds, the health plan may pay a lower cost to providers by using either Tiered or Value-based Networks.*

First, consider the case where condition (30) for Provider A is violated. By Proposition 6, in Value-based Networks the health plan pays Provider A less by using prospective payment than by using cost reimbursement. Now compare Provider A's gross profit margin $p_A - C(q_A, e_A)$ in Value-based Networks to its cost margin m_A in Tiered Networks. In the Appendix, we show that the latter has a smaller value if and only if

$$\widehat{x}^*[vq_A^* - C(q_A^*, e_A^*)] - [G(q_A^*) + H(e_A^*)] > \widehat{x}^*[vq_B^* - C(q_B^*, e_B^*)] - [G(q_B^*) + H(e_B^*)].$$

However, this inequality always holds because in the first best, q_A^* and e_A^* maximize $\widehat{x}^*[vq_A - C(q_A, e_A)] - [G(q_A) + H(e_A)]$. Therefore, Provider A earns a lower profit under cost reimbursement in Tiered Networks, compared to either cost reimbursement or prospective payment in Value-based Networks. Because Provider B also earns a lower profit in Tiered Networks, Value-based Networks are dominated by cost reimbursement in Tiered Networks in terms of implementation cost. The health plan's optimal choice is between cost reimbursement and prospective payment in Tiered Networks, and the tradeoff is fully characterized by Proposition 3.

When condition (30) for Provider A holds, the health plan pays less to Provider A by cost reimbursement rather than by prospective payment in Value-based Networks. We now use an example to illustrate that in this case, the health plan may pay a lower cost to providers by using either Tiered or Value-based Networks.

Example 1 Suppose that the fixed costs of quality and effort are quadratic, that is, let $G(q) = Gq^2$ and $H(e) = He^2$, where G and H are some positive parameters. Using (13) and (14), we can express the total cost margin, $m_A\widehat{x}^* + m_B(1 - \widehat{x}^*)$, in Tiered Networks as

$$\left[\frac{Gq_A^{*2} + He_A^{*2} - Gq_B^{*2} - He_B^{*2}}{\widehat{x}^* - 1/2}\right]\widehat{x}^* + \beta\left[\frac{Gq_B^{*2} + He_B^{*2}}{1 - \widehat{x}^*}\right](1 - \widehat{x}^*). \tag{31}$$

Similarly, using (28) and (29), we can express the total cost margin in Value-based Networks as

$$\left[\frac{(2/q_A^*)(Gq_A^{*2} + He_A^{*2})}{v/2\tau}\right]\widehat{x}^* + \beta\left[\frac{(2/q_B^*)(Gq_B^{*2} + He_B^{*2})}{v/2\tau}\right](1 - \widehat{x}^*). \tag{32}$$

Upon simplification, the value of (31) is smaller than the value of (32) if $\beta \leq 2$ and

$$\frac{1}{\widehat{x}^* - 1/2}(Gq_A^{*2} + He_A^{*2}) < \frac{1}{v/2\tau}(2/q_A^*)(Gq_A^{*2} + He_A^{*2}), \tag{33}$$

but the reverse can be true if inequality (33) fails to hold.[9]

In general, the comparison between the cost margins in Tiered and Value-based Networks depends on the values of $G(q)$, $H(e)$, and their first-order and second-order derivatives at the first best. The quadratic fixed costs in Example 1 allow us to simplify the comparison, and focus on the incentive properties of network design. On the left-hand side of (33), $\widehat{x}^* - 1/2$ is the demand response when Provider A raises quality from q_B^* to q_A^* in Tiered Networks, and $Gq_A^{*2} + He_A^{*2}$ is the total fixed cost of Provider A. On the right-hand side of (33), $\frac{v}{2\tau}$ is the demand response when Provider A raises q_A by one unit in Value-based Networks, whereas $(2/q_A^*)(Gq_A^{*2} + He_A^{*2})$ is the equivalent of the first-best marginal fixed costs under the quadratic costs assumption. Example 1 shows that when condition (30) for Provider A holds, neither Tiered nor

[9] Subtract (31) from (32) and rearrange terms, the value of (31) is smaller than the value of (32) if and only if $\left[\frac{2/q_A^*}{v/2\tau} - \frac{1}{\widehat{x}^* - 1/2}\right][Gq_A^{*2} + He_A^{*2}]\widehat{x}^* + \beta\left[\frac{2/q_B^*}{v/2\tau} - \frac{1}{\widehat{x}^* - 1/2}\left\{\frac{\beta(\widehat{x}^* - 1/2) - \widehat{x}^*}{\beta(1 - \widehat{x}^*)}\right\}\right][Gq_B^{*2} + He_B^{*2}](1 - \widehat{x}^*) > 0$. Because $q_A^* < q_B^*$ and $\frac{\beta(\widehat{x}^* - 1/2) - \widehat{x}^*}{\beta(1 - \widehat{x}^*)} \leq 1$ for any $\widehat{x}^* < 1$ if $\beta \leq 2$, the value of the first square-bracketed term is smaller than the value of the second square-bracketed term. Moreover, the value of the first square-bracketed term is positive if inequality (33) is satisfied.

Valued-based Network must dominate the other in terms of implementation cost. In the example, a network design is more likely to have a lower implementation cost if it can generate a stronger demand response.

6 Conclusion

In this paper, we have studied health plan designs. Our contribution is threefold. First, we introduce Tiered and Value-based Networks. A health plan uses one of these two disclosure methods to inform consumers about providers' qualities and costs, as well as to incentivize providers to choose qualities and cost efforts. Second, we pair each disclosure method with a payment mechanism, either cost reimbursement or prospective payment, for the implementation of efficient qualities and cost efforts. Third, we set consumer copayments to implement the efficient allocations of consumers across providers. We have derived conditions for each type of networks to implement the first best. Where the first best can be implemented, we compare the associated costs in different networks.

Our approach illustrates the multifaceted nature of health plan designs. The previous literature has contrasted provider incentives due to cost reimbursement and prospective payment. We view payment mechanism only as a component in the overall health plan design. Together, payment mechanisms for providers and information mechanism for consumers determine how providers choose qualities and cost effort, and how consumers choose providers. We show how health plans can exploit this interaction. In fact, we show how health plans may implement first best in various ways, which may require different implementation costs.

We have looked at payment policies and information-based health plans with a broad perspective. While financial incentives are certainly important, our innovation is to supplement provider financial incentives with information disclosure to consumers. The simplistic way of disclosing all relevant information to consumers misses the point. Optimal network design considers optimal information disclosure. Both Tiered and Value-based Networks add to the power of financial incentives. Information-based policies may prove powerful in many aspects of health care deliveries.

Appendix

Proof of Lemma 1: 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 are 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 provider 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 perfect-Bayesian 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

$$\begin{aligned} m_A &\geq \frac{G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2} && (34) \\ m_B &\leq \frac{\beta[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} && (35) \\ m_A &\geq \frac{[G(q_A^*) + H(e_A^*)]}{\widehat{x}^*} && (36) \\ m_B &\geq \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}. && (37) \end{aligned}$$

A sufficiently big value of m_A can always be chosen to satisfy both (34) and (36), so the first best is implementable if and only if there is an m_B satisfying both (35) and (37). This is equivalent to

$$\frac{\beta[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} \geq \frac{\beta[G(q_B^*) + H(e_B^*)]}{1 - \widehat{x}^*}, \tag{38}$$

which simplifies to (12) in the Proposition.

Proof of Corollary 2: The minimum value for m_B has already been derived just before the presentation of the Corollary.

Consider Provider A. The value of m_A must satisfy both (34) and (36) (see the proof of Proposition 1). The minimum value for m_A must be the larger of the right-hand side expressions of (34) and (36). We now show that the right-hand side expression in (34) 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 numerator. That is, the right-hand side of (34) is larger than the right-hand side of (36) if and only if $\dfrac{G(q_A^*) + H(e_A^*)}{G(q_B^*) + H(e_B^*)} > 2\widehat{x}^*$.

Now by (12), it is sufficient to show that $\dfrac{1}{2(1-\widehat{x}^*)} > 2\widehat{x}^*$. For this, we calculate

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

Hence, the lowest possible first-best equilibrium m_A is the one in (13).

Proof of Proposition 2: Rearranging inequalities (15) to (18), we get

$$p_A \geq \frac{C(q_A^*, e_A^*)\widehat{x}^* - C(q_B^*, e_B^*)/2 + G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2} \tag{39}$$

$$p_B \leq \frac{C(q_A^*, e_A^*)/2 - C(q_B^*, e_B^*)(1-\widehat{x}^*) + \beta[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2} \tag{40}$$

$$p_A \geq C(q_A^*, e_A^*) + \frac{[G(q_A^*) + H(e_A^*)]}{\widehat{x}^*} \tag{41}$$

$$p_B \geq C(q_B^*, e_B^*) + \frac{\beta[G(q_B^*) + H(e_B^*)]}{1-\widehat{x}^*}. \tag{42}$$

Conditions (39) and (41) can be satisfied by setting p_A to a sufficiently large value. Therefore, the first best is implementable only if there exists p_B that satisfies (40) and (42). This is equivalent to

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

which can be rewritten as

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

Next, we note that inequality (12) in Proposition 1 follows from inequality (38) in the proof of that Proposition. By inspection, we only have to simplify the second term on each side of the above. Indeed, after the simplification, we obtain (19).

Proof of Corollary 3: We obtain (23) by setting expression (18) as an equality.

For Provider A, the minimum value of p_A is the larger of the right-hand side expressions of (15) and (17). The difference between the two expressions is

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

This expression is positive if and only if condition (20) is satisfied. When condition (20) holds, the minimum p_A must satisfy (15) as an equality. When condition (20) fails, the minimum p_A must satisfy (17) as an equality. We thereby obtain expressions (21) and (22) for the minimum p_A.

Proof of Proposition 3: We have shown that $m_B + C(q_B^*, e_B^*) = p_B$ before the presentation of the Proposition. We now compare the minimum values of $m_A + C(q_A^*, e_A^*)$ and p_A. There are two cases. Case 1: The prospective price is given by (21) in Corollary 3 (which is equivalent to inequality (20) satisfied). Using expressions (13) and (21), we have

$$\begin{aligned}
& p_A - [m_A + C(q_A^*, e_A^*)] \\
= & \left[\frac{C(q_A^*, e_A^*)\widehat{x}^* - C(q_B^*, e_B^*)/2 + G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2}\right] \\
& \quad - \left[\frac{G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)}{\widehat{x}^* - 1/2}\right] - C(q_A^*, e_A^*) \\
= & \frac{C(q_A^*, e_A^*) - C(q_B^*, e_B^*)}{2\widehat{x}^* - 1},
\end{aligned}$$

which is positive if and only if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$.

Case 2: The prospective price is given by (22), which is equivalent to inequality (20) failing to hold:

$$C(q_A^*, e_A^*) + \frac{G(q_A^*) + H(e_A^*)}{\widehat{x}^*} < C(q_B^*, e_B^*) + \frac{G(q_B^*) + H(e_B^*)}{1/2},$$

which simplifies to

$$2\widehat{x}^*[G(q_B^*) + H(e_B^*)] + \widehat{x}^*[C(q_B^*, e_B^*) - C(q_A^*, e_A^*)] > G(q_A^*) + H(e_A^*). \tag{43}$$

In this case, Provider A makes a positive profit under cost reimbursement and zero profit under prospective payment. Hence we have $p_A < m_A + C(q_A^*, e_A^*)$. Now from Proposition 2, first-best implementation requires (19). This condition and (43) can be satisfied simultaneously only if

$$C(q_B^*, e_B^*) - C(q_A^*, e_A^*) > \frac{(1 - 2\widehat{x}^*)^2}{(2\widehat{x}^* - 1/\beta)(1 - \widehat{x}^*)}[G(q_B^*) + H(e_B^*)] > 0,$$

where the last inequality follows from $\widehat{x}^* > 1/2$ and $\beta > 1$.

We conclude that $p_A \geq m_A + C(q_A^*, e_A^*)$ if and only if $C(q_A^*, e_A^*) \geq C(q_B^*, e_B^*)$.

Proof of Lemma 4: Suppose that in equilibrium Provider A's market share is $\widehat{x}$. Let equilibrium (q_A, e_A) achieve index $\theta_A q_A +$ $(1 - \theta_A)e_A$. Next, suppose that Provider A deviates to any (q'_A, e'_A) where $\theta_A q'_A +$

$(1-\theta_A)e'_A = \theta_A q_A + (1-\theta_A)e_A$. Consumers must continue to believe that the quality and cost effort remain at q_A and e_A. Provider A's demand remains at $\widehat{x}$. Moreover, $m_A\widehat{x} - [G(q'_A) + H(e'_A)] \leq m_A\widehat{x} - [G(q_A) + H(e_A)]$, with a strict inequality when $(q'_A, e'_A) \neq (q_A, e_A)$ due to the strict convexity of G and H. Hence (q'_A, e'_A) is not a profitable deviation and (q_A, e_A) has the property as stated in the Lemma. The proof for equilibrium (q_B, e_B) follows symmetrically. Finally, (25) are the first-order conditions of the two profit maximization programs in the Lemma. Equilibrium (q_i, e_i), $i = A, B$, must satisfy (25).

Proof of Proposition 4: Suppose that the quality and cost effort of Provider B is (q_B^*, e_B^*). Consider the profit-maximization problem of Provider A. By Corollary 4, the first-best quality and effort satisfy $\dfrac{G'(q_A^*)}{H'(e_A^*)} = \dfrac{\theta_A}{1-\theta_A}$ and hence can be Provider A's optimal choice. By Lemma 4, we use (25) to define e_A as a function of q_A, say $e_A(q_A)$. Then we calculate its derivative:

$$e'_A(q_A) = \frac{1-\theta_A}{\theta_A}\frac{G''(q_A)}{H''(e_A)} > 0. \tag{44}$$

We then rewrite Provider A's constrained profit maximization program as

$$\max_{q_A}\ m_A\left[\frac{1}{2} + \frac{v(q_A - q_B) - (s_A - s_B)}{2\tau}\right] - [G(q_A) + H(e_A(q_A))].$$

The first-order condition is

$$m_A\frac{v}{2\tau} = G'(q_A) + H'(e_A)e'_A(q_A). \tag{45}$$

The margin m_A for first-best implementation must satisfy (45) at $q_A = q_A^*$ and $q_B = q_B^*$. This margin is unique because both $G'(q_A)$ and $H'(e_A)e'_A(q_A)$ are increasing in q_A. We obtain the expression for m_A in the Proposition by using equations (25), (44), and the identity $d\ln f = f'/f$ to simplify (45). The derivation of m_B follows the same steps. Given the two margins in the Proposition, no unilateral deviation from the first best is profitable. Moreover, the consumer demand for services at Provider A is

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

Hence, the first best is implemented as a perfect-Bayesian equilibrium.

Proof of Lemma 5: Suppose in equilibrium Provider A serves $\widehat{x}$ of consumers. Let the equilibrium prospective price and index weight be p_A, and θ_A, respectively. Equilibrium quality and cost effort (q_A, e_A)

must solve

$$\max_{q'_A, e'_A} \; [p_A - C(q'_A, e'_A)]\widehat{x} - [G(q'_A) + H(e'_A)] \quad \text{s.t.} \;\; \theta_A q'_A + (1-\theta_A)e'_A = \theta_A q_A + (1-\theta_A)e_A.$$

After simplifying the first-order conditions, and setting $q'_A = q_A$, $q'_B = q_B$, we have

$$-\widehat{x}C_e(q_A, e_A) = H'(e_A) - \frac{(1-\theta_A)}{\theta_A[G'(q_A) + \widehat{x}C_q(q_A, e_A)]}. \tag{46}$$

Let $\widehat{x} = \widehat{x}^*$ and $q_A = q_A^*$. Condition (46) is identical to (4), which implicitly defines e_A^*, only if $\theta_A = 1$.

The proof that the first best is implementable only if $\theta_B = 1$ follows the same steps.

Proof of Proposition 5: Let the quality and cost effort of Provider B be (q_B^*, e_B^*). By Lemma 5, $\theta_A = 1$ and hence consumers perfectly learn q_A from the value index. Provider A chooses q_A and e_A to maximize

$$[p_A - C(q_A, e_A)]\left[\frac{1}{2} + \frac{v(q_A - q_B) - (s_A - s_B)}{2\tau}\right] - [G(q_A) + H(e_A)].$$

The first-order conditions with respect to q_A and e_A are

$$[p_A - C(q_A, e_A)]\frac{v}{2\tau} - C_q(q_A, e_A)\widehat{x} \;=\; G'(q_A) \tag{47}$$

$$-C_e(q_A, e_A)\widehat{x} \;=\; H'(e_A), \tag{48}$$

where $\widehat{x} \equiv \left[\frac{1}{2} + \frac{v(q_A - q_B) - (s_A - s_B)}{2\tau}\right]$. Conditions (4) and (48) are identical at $q_A = q_A^*$ and $e_A = e_A^*$. To obtain the optimal prospective price p_A, consider conditions (2) and (47). They are identical if and only if

$$\widehat{x}^*[v - C_q(q_A^*, e_A^*)] = [p_A - C(q_A^*, e_A^*)]\frac{v}{2\tau} - C_q(q_A^*, e_A^*)\widehat{x}^*.$$

Simplifying further, we get

$$\begin{aligned} p_A &= 2\tau\widehat{x}^* + C(q_A^*, e_A^*) \\ &= 2\tau\left[\frac{1}{2} + \frac{v(q_A^* - q_B^*) - [C(q_A^*, e_A^*) - C(q_B^*, e_B^*)]}{2\tau}\right] + C(q_A^*, e_A^*) \\ &= \tau + C(q_B^*, e_B^*) + v(q_A^* - q_B^*). \end{aligned}$$

Given this value of p_A, (q_A^*, e_A^*) is the unique solution to (47) and (48). The optimal p_B is obtained by the same steps. Finally, the demand for Provider A's services is

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

This completes the first-best implementation.

Proof of Proposition 6: First, in the proof of Proposition 4 we have shown that $e'_A(q_A^*) = \dfrac{d\ln G'(q_A^*)}{d\ln H'(e_A^*)}$. Using this identity, we can rewrite the first-order condition (45) under cost reimbursement as

$$\frac{m_A \dfrac{v}{2\tau} - G'(q_A^*)}{H'(e_A^*)} = \frac{d\ln G'(q_A^*)}{d\ln H'(e_A^*)}. \tag{49}$$

Second, combining first-order conditions (47) and (48) under prospective payment, we have

$$\frac{[p_A - C(q_A^*, e_A^*)]\dfrac{v}{2\tau} - G'(q_A^*)}{H'(e_A^*)} = -\frac{C_q(q_A^*, e_A^*)}{C_e(q_A^*, e_A^*)}. \tag{50}$$

By inspecting (49) and (50), we conclude that $p_A > m_A + C(q_A, e_A)$ if and only if $-\dfrac{C_q(q_A^*, e_A^*)}{C_e(q_A^*, e_A^*)} > \dfrac{d\ln G'(q_A^*)}{d\ln H'(e_A^*)}$. The proof that $p_B > m_B + C(q_B^*, e_B^*)$ if and only if $-\dfrac{C_q(q_B^*, e_B^*)}{C_e(q_B^*, e_B^*)} > \dfrac{d\ln G'(q_B^*)}{d\ln H'(e_B^*)}$ follows the same steps.

Proof of Proposition 7: Suppose that condition (30) for Provider A is violated. We now show that the health plan incurs a lower payment to providers by using Tiered Networks. First, because Provider B's nonnegative constraint is only binding in Tiered Networks, the health plan always minimizes the payment to Provider B by using Tiered Networks, irrespective of the choice of the payment system.

Next, consider Provider A. By Proposition 6, the health plan minimizes the payment to Provider A by using prospective payment in Value-based Networks if the Provider's cost functions violate condition (30). Now compare Provider A's gross profit margin under prospective payment in Value-based Networks, $p_A - C(q_A, e_A)$, to its cost margin under cost reimbursement in Tiered Networks, m_A. By Corollary 2 and Proposition 5, the value of $p_A - C(q_A, e_A)$ is bigger if and only if

$$\tau + v(q_A^* - q_B^*) - [C(q_A^*, e_A^*) - C(q_B^*, e_B^*)] > \frac{[G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)]}{\widehat{x}^* - 1/2}.$$

Using the definition of $\widehat{x}^*$ in (6) and noting that $\widehat{x}^* - 1/2 = \dfrac{v(q_A^* - q_B^*) - [C(q_A^*, e_A^*) - C(q_B^*, e_B^*)]}{2\tau}$, we can rewrite the inequality as

$$\widehat{x}^*\{v(q_A^* - q_B^*) - [C(q_A^*, e_A^*) - C(q_B^*, e_B^*)]\} > G(q_A^*) + H(e_A^*) - G(q_B^*) - H(e_B^*)$$

or

$$\widehat{x}^*[vq_A^* - C(q_A^*, e_A^*)] - [G(q_A^*) + H(e_A^*)] > \widehat{x}^*[vq_B^* - C(q_B^*, e_B^*)] - [G(q_B^*) + H(e_B^*)]. \tag{51}$$

By Lemma 1, the values of q_A^* and e_A^* are chosen to maximize $\widehat{x}^*[vq_A - C(q_A, e_A)] - [G(q_A) + H(e_A)]$ and $(q_A^*, e_A^*) \neq (q_B^*, e_B^*)$. Because of the strict convexity of C, G, and H, inequality (51) holds. Hence the health plan incurs a lower payment to Provider A by using cost reimbursement in Tiered Networks, compared to using either cost reimbursement or prospective payment in Value-based Networks. This completes the proof that the health plan minimizes the payment to providers by using Tiered Networks when condition (30) for Provider A is violated.

Finally, in the discussion after the statement of the Proposition, we show by an example that if condition (30) for Provider A is satisfied, the health plan may minimize the payment to providers by using either Tiered or Value-based Networks. This completes the proof of the Proposition.

References

Alteras, T., and S. Silow-Carroll. "Value-Driven Health Care Purchasing: Case Study of the Massachusetts Group Insurance Commission." The Commonwealth Fund pub. no. 1053 (2007).

Baicker, K., S. Mullainathan, and J. Schwartzstein. "Behavioral Hazard in Health Insurance." Working paper (2012), Harvard University.

Beitia, A. "Hospital Quality Choice and Market Structure in a Regulated Duopoly." *Journal of Health Economics* 22, no. 6 (2003): 1011-1036.

Christianson, J.B., K.M. Volmar, J. Alexander, and D.P. Scanlon. "A Report Card on Provider Report Cards: Current Status of the Health Care Transparency Movement." *Journal of General Internal Medicine* 25 (2010): 1235-1241.

Dranove, D., and G.Z. Jin. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48, no. 4 (2010): 935-963.

Dranove, D., and A. Sfekas. "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards." *Journal of Health Economics* 27, no. 5 (2008): 1201-1207.

Kaiser Family Foundation and Health Services & Educational Trust (2013). Employer Health Benefits: 2013 Annual Survey, from http://kaiserfamilyfoundation.files.wordpress.com/2013/08/8465-employer-health-benefits-20131.pdf

Léger, P.T. "Physician Payment Mechanisms." in *Financing Health Care: New Ideas for a Changing Society*, edited by Lu, M. and E. Jonsson, 149-176, Wiley-VCH (2008).

Levaggi, R. "Hospital Health Care: Pricing and Quality Control in a Spatial Model with Asymmetry of Information." *International Journal of Health Care Finance and Economics* 5, no. 4 (2005): 327-349.

Ma, C.A. "Health Care Payment Systems: Cost and Quality Incentives." *Journal of Economics & Management Strategy* 3, no 1 (1994): 93-112.

Ma, C.A. and H.Y. Mak. "Information Disclosure and the Equivalence of Prospective Payment and Cost Reimbursement." Working paper (2013), Boston University.

McGuire, T.G. "Demand for Health Insurance." in *Handbook of Health Economics*: vol. 2, edited by Pauly, M.V., T.G. McGuire, and P.P. Barros, 317-396, Elsevier (2011).

Mehrotra, A., P.S. Hussey, A. Milstein, and J.H. Hibbard. "Consumers' and Providers' Responses to Public Cost Reports, and How to Raise the Likelihood of Achieving Desired Results." *Health Affairs* 31, no. 4 (2012): 843-851.

Milgrom, P., and C. Shannon. "Monotone comparative statics." *Econometrica* 62, no. 1 (1994): 157-180.

Pauly, M.V., and F.E. Blavin. "Moral Hazard in Insurance, Value-Based Cost Sharing, and the Benefits of Blissful Ignorance." *Journal of Health Economics* 27, no. 6 (2008): 1407-1417.

Robinson, J. "Hospital Tiers in Health Insurance: Balancing Consumer Choice with Financial Incentives." *Health Affairs* 22, no. 3; SUPP (2003): W3-135.

Spence, M.A. "Monopoly, Quality, and Regulation." *Bell Journal of Economics* 6, no. 2 (1975): 417-429

Thomson, S., L. Schang, and M.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, no. 4 (2013): 704-712.