Bundled Payments For Healthcare Services: Proposer Selection and Information Sharing

Diwakar Gupta
Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota 55455, guptad@umn.edu

Mili Mehrotra
Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455, milim@umn.edu

The Centers for Medicare and Medicaid Services (CMS) has introduced a “bundled payments for care improvement” (BPCI) initiative. Each bundle pertains to a specific medical condition, a set of linked services, and a length of time referred to as an episode of care. Proposers choose bundles, design service chains, and propose target values of quality metrics and payments per episode. Expert panels evaluate proposals based on CMS-announced relative weights, but there is no limit on the number of proposers that may be selected. Moreover, there is no minimum score that will guarantee selection, which makes selection uncertain for proposers. We develop normative models for the parameter selection problems faced by potential proposers within the CMS’ proposal selection process. Proposers have private information about their costs of achieving different quality targets, which determine their equilibrium responses. We show that an optimal strategy for CMS, under its current approach, may be to either announce a fixed threshold or keep the selection process uncertain, depending on market characteristics. We also formulate and solve the proposer selection problem as a constrained mechanism design problem, which reveals that CMS’ current approach is not optimal. We present policy guidelines for government agencies pursuing bundled payment innovations.

Subject classifications: games/group decisions: bidding/auction; government: agencies; healthcare.

Area of review: Policy Modeling and Public Sector OR.

History: Received October 2014; revisions received February 2015, April 2015; accepted May 2015.

1. Introduction

U.S. healthcare expenditures are projected to be about $3.2 trillion in 2015 (CMS 2015). In addition to its high cost, the U.S. healthcare delivery system is frequently criticized for being a fragmented system with fee-for-service payments to medical professionals that result in uncoordinated and duplicate services across the outpatient, inpatient, and long-term care continuum. Consequently, payers have begun to bundle payments as a means to improve accountability and health outcomes, and reduce expenditures (Larkin 2010, Markovich 2012, CMS 2013). In this paper, we focus on the bundled payments for care improvement (BPCI) initiative by the Centers for Medicare and Medicaid Services (CMS). CMS’ Medicare and Medicaid programs together accounted for about 36% of national healthcare expenditures in 2012, whereas all private insurers combined covered about 33% (CMS 2014).

BPCI invites proposers to define bundles of services related to a particular medical condition such as a major joint replacement and submit proposals that include target payments per episode, detailed service designs, quality targets, and plans for monitoring and dealing with outliers (Levy 2012). Proposers may be either physician groups, acute care hospitals, integrated health systems, physician-hospital organizations, or conveners. CMS does not specify a minimum score (a fixed threshold) above which selection will be assured. Panels of experts score proposals according to CMS-specified weights and decide which proposals will be selected, making the selection process uncertain. Proposers do not compete for selection because there is no limit on the number of proposals that may be selected. However, proposers compete to attract beneficiaries. We refer to CMS’ proposal selection process as the BPCI mechanism.

Our goals in this paper are to (1) develop normative models to evaluate the BPCI mechanism and (2) study properties of a constrained optimal mechanism. Our efforts help answer the following questions. How should rational proposers respond to CMS’ request for applications (RFA)? Should CMS, which plays the role of a social planner, either reduce or eliminate uncertainty in the selection process via information sharing? Uncertainty is reduced when CMS specifies detailed characteristics of bundles and proposers that are more likely to be selected. To eliminate uncertainty completely, CMS may announce a fixed threshold as its selection criterion. A relevant question in the latter case would be that if CMS were to announce a fixed threshold, what should be its optimal value? What would be an optimal mechanism, given attribute weights used by the BPCI mechanism? How can agencies like CMS improve health services’ procurement in future contracts?
Next, we describe the BPCI mechanism, which forms the basis for our modeling choices (see CMS 2013 for additional details). A bundle consists of the following proposer-selected parameters: (1) the medical condition, (2) linked services, and (3) the bookends of an episode of care. Medical conditions have been classified into diagnosis-related groups (DRGs)—see CMS (2013) for a list of DRGs that have been selected for the BPCI program. Different proposers may define bundles differently, even for the same DRG, depending on which services are linked and the length of the episode of care. For example, one proposer’s knee-joint replacement bundle may include all related surgical, anesthesia, radiology, and imaging services, discharge planning, and home care, as well as the implant, pharmaceuticals, and surgical supplies, whereas some of these components may be excluded in another proposer’s bundle. Similarly, in a proposal that links acute and post acute care, an episode of care may begin with a hospital admission and end either 30, 60, or 90 days after hospital discharge. An enrollee is a beneficiary who triggers an episode by utilizing a bundled service at a BPCI provider. Beneficiaries may receive either all or a portion of services included in the bundle from any provider. BPCI providers may neither cherry-pick nor lock in enrollees.

CMS has grouped linked services and payment modalities into four models (see additional background information provided in Appendix A. In this paper, we analyze the BPCI mechanism in which proposers are paid prospectively. That is, proposers after being selected for particular medical conditions (DRGs) and episode bookends receive single payments equal to their proposed targets, keep all gains if actual expenditures are smaller, and incur a loss otherwise. The individual service providers comprising the proposer do not bill separately for each service.

Proposals may be characterized by two scalar parameters: a target payment per episode and a proposed quality score. Consistent with CMS’ current practice, both are evaluated relative to historical averages. For example, a proposal is evaluated in terms of the discount it offers relative to the sum of the historical average payments made to clinics, hospitals, doctors, and long-term care facilities that constitute the proposer. A minimum discount is required by CMS. Similarly, the quality score is the net improvement in the weighted aggregate score across all metrics not related to target payments that are of interest to the payer.

We model CMS’ current practice of measuring cost and quality relative to historical averages. One may wonder why CMS evaluates proposals relative to historical benchmarks. Reasons for doing so may include budget constraints, differences in the severity mix of patients served by different providers, significant variability in payments and quality metrics across providers for services related to the same DRG, and greater motivation for proposers to innovate if the benchmark is proposer specific. CMS’ scheme rewards those proposers that achieve higher quality at lower cost, while adjusting target payments by historical severity mix.

Alternate schemes are possible. Their analysis is beyond the scope of this paper.

Each proposer must credibly demonstrate its ability to achieve the proposed quality score by appropriately designing its service chain. CMS requires regular reporting of quality metrics and utilization rates, and may terminate the contract if the awardee fails to meet proposed standards. We assume that the financial loss from termination is sufficiently large that proposers choose achievable quality scores. We further assume that proposers can estimate their optimal cost functions, i.e., the relationships between each target quality score and the corresponding minimum-cost service chain design. Proposers’ costs are private information. They include the direct cost of providing services, the cost of coordination across providers, and the cost of leakage, i.e., payments made to those providers that are not part of the proposers’ networks and whose services are deemed part of the bundle. Private information also includes each proposer’s historical average payments for the proposed bundle. All private information is encapsulated in a proposer type. Because new service chains form in response to the BPCI program, proposers do not know each other’s types. The payer does not know proposers’ types but knows historical payments made to each provider.

Using guidelines provided to expert panels, proposers have the ability to score their proposals. However, they do not know which panel will review their proposals and whether that panel will recommend their proposals for selection. Therefore, each proposer assumes that its proposal has a probability of being selected, which is independent of how other proposers score their own proposals. The latter is justified by the fact that proposers do not compete with each other to be selected and different proposals may be evaluated by different expert panels. However, the number of enrollees that a proposer can attract does depend on which other proposers are selected and their proposed scores (Pearce and Harris 2010).

The remainder of the paper is organized as follows. Section 2 summarizes the relevant literature. In §3, we describe a proposer’s demand function, set up the proposer’s and the payer’s problems, and provide our preliminary analysis for obtaining the equilibrium responses of proposers. Recall that our goal in this paper is to understand the extent to which CMS should share information. In §4, we illustrate that it is beneficial for the payer to have selection uncertainty when there is either information asymmetry or competition or both. To this end, we consider two special cases: single proposer with multiple types (§4.1.1) and multiple proposers of the same type (§4.1.2). In §4.2, we generalize our insights to settings with multiple proposers and multiple types. We present and analyze the optimal mechanism in §5. Finally, §6 summarizes our insights, provides future research directions, and concludes the paper.
2. Literature Review

Because the BPCI program, launched in 2011, is relatively new, there is limited literature on related systemic and operational issues (see, e.g., Rana and Bozic 2014). To the best of our knowledge, ours is the first attempt to evaluate the BPCI mechanism via mathematical models and consider the effect of information sharing. The BPCI mechanism possesses unique features that make our model and results significantly different from those discussed in the literature on health services, health economics, and procurement.

Hussey et al. (2012), summarize evidence from 58 studies (both United States and international) concerning previous attempts to bundle payments to healthcare providers. Most of these studies (57 out of 58) were descriptive or observational and 16 out of 20 bundled payment interventions focused on a single institutional provider. None concerned normative models, which is our focus. The authors found that there was weak but consistent evidence that bundled payment programs reduced costs without major effects on quality. Reductions in expenditures and utilization of services relative to those under the usual payment mechanism were under 10% in many cases. All previous CMS-sponsored projects reviewed in Hussey et al. (2012) were demonstration projects. In contrast, BPCI is intended for widespread adoption and to spur innovation in care delivery.

Dobson and DaVanzo (2012) used beneficiary-level claims data to analyze the effects of bundled payments on various types of providers with the help of descriptive statistics and multivariate regression models. They identify beneficiary and provider characteristics that affect Medicare episode payments. They also identify patient pathways, readmissions, and other episode characteristics that a provider needs to control to effectively manage costs. Examples of other themes pursued in health services and health economics literatures include (1) comparisons of bundled payments and other payment innovations (McClellan 2011, Suskind and Clemens 2014), (2) potential for cost savings from bundled payment programs (Mechanic and Tompkins 2012, Miller et al. 2011, Cutler and Ghosh 2012), (3) potential for quality improvement because of improved postacute care (Mead et al. 2014), and (4) opinion pieces that provide advice to service providers (Altman 2012) and CMS (Delisle 2013, Rosen et al. 2013). The vast majority of authors base their advice and conclusions on earlier attempts at bundling payments, which were fundamentally different from the BPCI program. For example, in the BPCI program, proposers select target payments so long as they do not exceed CMS specified maximum, and there is no pay for performance.

Advice offered to CMS consists of the need to pay attention to different components of bundles, e.g., adjustment for riskier patients and sending consistent signals to service providers (Altman 2012, Rosen et al. 2013). The BPCI program responds to these concerns by allowing proposers to choose target payments and to choose from a menu of payment models in which different amounts of risks are shared with the proposers. In particular, we focus in this paper on the prospective payment model in which the proposers assume all of the risk and keep all of the benefit from cost savings. Also, proposers are permitted to form networks of service providers both to spread risk across more patients and to create larger and integrated service chains.

A second type of advice is offered to proposers by reminding them to price risk appropriately and to focus attention on practice areas that have the potential to reduce costs (e.g., Miller et al. 2011, Sood et al. 2011). Proposers in our model are assumed to be risk neutral. However, risk could be priced and added to a proposer’s cost as the price of an option that eliminates its loss under different shared savings contracts, as illustrated in Friedberg et al. (2013). That is, our models indirectly account for different levels of risk faced by different types of proposers, depending on their size and patient mix, by modeling costs as proposer-dependent and private information.

Our work differs from the papers discussed above due to its focus on analyzing the proposal-selection mechanism and information sharing with the help of mathematical models while considering the effect of competition among proposers to attract beneficiaries.

Turning next to the supplier-selection problem in the procurement literature, the problem considered in our paper falls under the principal-agent framework with multiple agents. In such problems, agents have private information, and their actions after selection may be hidden. The principal is therefore concerned with (1) the selection of the best agents and (2) the control of their actions after selection. Auction theory focuses primarily on the former and agency theory focuses primarily on the latter (Seshadri 1995). Auction theory considers two types of models: (1) competitive bidding for selecting and allocating procurement quantity to a subset of suppliers (Dasgupta and Spulber 1990, Seshadri et al. 1991) and (2) collusive bidding where everyone who bids may receive a portion of the quantity (Wilson 1979, Anton and Yao 1989). Because agents’ actions are hidden, agency theory considers the design of either incentives that are based on the output/costs of the agents (Laffont and Martimort 2002) or tournaments/contests with agents earning rewards according to their ordinal ranks (Lazear and Rosen 1981, Nalebuff and Stiglitz 1983). Excellent surveys on auctions, bids, and the use of incentives can be found in Engelbrecht-Wiggans et al. (1983), McAfee and McMillan (1987), and Laffont and Tirole (1993).

In the healthcare literature, there are papers that have utilized agency theory to design incentives to reduce costs (Shleifer 1985, Newhouse 1996) or to improve health outcomes (Fuloria and Zenios 2001, Lee and Zenios 2012). In contrast to these papers, a unique feature of the BPCI mechanism, and therefore of our model, is that it focuses on both reducing cost and improving health outcomes. Furthermore, the regular reporting requirement and the threat...
of disqualification are assumed to be sufficient in our setting to ensure that proposers offer achievable quality scores and there is no need to control proposers’ actions after selection. That being the case, our work is closer to papers that model auctions with multiple agents’ selection.

Previous works on modeling auctions can be classified according to the assumptions underlying the models. The most common assumptions are (1) known number of bidders, (2) independent private values (IPV), (3) symmetric and risk-neutral bidders, (4) risk-neutral buyer, and (5) the assumption that the final payment to the winner is either firm, or includes either a penalty or a bonus that is a linear function of the bid price. We also make these assumptions for tractability reasons. However, our paper differs from the existing literature in a number of key features and results. First, the proposers in our model do not compete for selection and there is no a priori limit on the number of proposers that may be selected. Second, the principal does not allocate demand to selected proposers, but rather, each selected proposer’s demand and profit is determined by how the beneficiaries respond to its offer. Third, we consider randomization, which affects bidders’ beliefs about selection uncertainty. That is, the level of uncertainty is endogenous.

There are several papers in the economics literature on adverse selection that compare deterministic and stochastic mechanisms in principal-agent models and identify cases in which the optimal mechanism may have randomness. For example, three such cases are (i) when agents’ utility functions belong to different risk aversion classes (Arnott and Stiglitz 1988), (ii) when agents have multidimensional type space (Baron and Myerson 1982), or (iii) when nonmonotonic allocation schedules are desirable (Straus 2006).

Compared to this literature, our setting does not have the three properties mentioned above. The proposers in our setting have identical linear utility functions and their type space can be reduced to a single dimension. However, we still observe that randomization is desirable in a number of cases. The key difference is that unlike the models discussed in the literature, the principal (the payer) in our setting can not allocate payoffs (demand) to agents (proposers). Proposers compete for beneficiaries and the demand allocation is an outcome of the Bayesian Nash game among proposers. We show in §5 that an uncertain mechanism is not optimal when there is no competition or the impact of competition is not significant on a proposer’s demand. That is, the competition among proposers to attract beneficiaries is the distinguishing feature in our setting that explains why randomization is desirable. This feature is not present in previous papers that consider randomization.

Among recent works, Giebe and Schweinzer (2015) and Athey and Nekipelov (2011) present symmetric scoring-auction models (for different applications) with quality and cost components in which (exogenous) selection uncertainty arises either from the buyer’s inability to assess bidder’s quality or from the fact that quality scores vary over time. After accounting for this uncertainty, either a single (Giebe and Schweinzer 2015) or a fixed number of winners (Athey and Nekipelov 2011) is selected. In contrast, we assume that CMS can reliably assess the quality of each proposer’s bundle. CMS chooses the level of information sharing, which affects bidders’ beliefs about selection uncertainty. That is, the level of uncertainty is endogenous. Also, CMS does not a priori announce how many winners it will select.

3. BPCI Mechanism: Model Formulation and Preliminary Results

Our model uses the following notation: $n$ for the number of potential proposers, $q$ for quality, $b$ for bundled payment, $t$ for type, $s_q$ for quality score, and superscript $h$ for historical values. We use uppercase letters to denote random variables, sets, and some functions, and overlines and underlines to denote either the upper and lower bounds of parameters or the limits of the range of possible values of decision variables. Increasing (decreasing) means nondecreasing (nonincreasing). We summarize notation in the electronic companion (available as supplemental material at http://dx.doi.org/10.1287/opre.2015.1403), Tables EC.1–EC.2.

A proposal is characterized by the proposed target payment ($b$) and quality score ($s_q$). Recall that the quality score ($s_q$) is the net improvement in the weighted aggregate score across all metrics not related to target payments that are of interest to the payer. A proposer’s private information, characterized by its type $t \in [\underline{t}, \bar{t}]$, consists of a function $c(s_q, t)$ that determines its average minimum cost of achieving a quality score $s_q$, its historical average payment for the proposed bundle after deducting mandatory discount required by the payer, which we denote by $b_h(t)$, and its estimated leakage, denoted by $\epsilon_t \geq 0$. All three quantities are differentiable in $t$ and specified on a per-episode basis. In this section, we assume that there is a continuum of proposer types in the range $[\underline{t}, \bar{t}]$. We introduce special cases with discrete types in §4.2. Some of the technical assumptions introduced later in this section are not needed for models with discrete types. In all cases, we assume symmetry, i.e., each proposer believes that its type is an independent random draw from a type distribution $F(\cdot)$. Proposers do not know other proposers’ types and the payer does not know any proposer’s type, but $F(\cdot)$ is common knowledge.

The sequence of events is as follows. The payer, acting as a social planner, announces a relative weight $\gamma > 0$ for quality. That is, $\gamma$ reflects a relative quality weight that all CMS beneficiaries collectively would assign to the quality attribute. We refer to $\gamma$ as the relative weight because we normalize the weight assigned to the offered discount to be equal to 1. Proposals are scored according to the rule: $r =$
\[ \gamma \cdot s_q + b^i(t) - b \] where \((s_q, b)\) with \(s_q \geq 0\) and \(0 \leq b \leq b^i(t)\) are the parameters of a proposal. We assume that proposers can score their own proposals, i.e., they know \(r\).

Two variants are modeled. If the payer announces a fixed threshold \(r^0\), then all proposals that achieve a score of at least \(r^0\) are selected and the rest are declined. Therefore, proposers who cannot achieve \(r^0\) do not propose. If the selection mechanism is uncertain, then expert panels recommend which proposals should be selected for the BPCI program. Each proposer believes that its proposal with score \(r \in [r, \bar{r}]\) has a probability \(G(r)\) of being selected, where \(\bar{r} \geq 0\) is the minimum acceptable and \(\bar{r} \geq r\) is the maximum achievable score. Proposers’ beliefs are affected by the information shared by the payer.

Consider the proposal-parameter selection problem faced by a type-\(t\) proposer. We refer to this proposer as the tagged proposer and omit proposer index and proposer type for sake of notational simplicity when describing its actions. The tagged proposer needs to select equilibrium parameters \((s_q, b)\) that maximize its individual profit, knowing that its demand will depend on the parameters selected by other proposers. Given \(G(r)\), the probability of being selected upon proposing score \(r\), the tagged proposer’s expected profit margin is \((b - c(s_q, t) - \epsilon_j)G(r) = (b - c(s_q, t) - \epsilon_j)\).\(\bar{G}(\gamma \cdot s_q + b^i(t) - b)\) per enrollee. Conditional upon being selected, the tagged proposer anticipates its demand to be \(d(s_q)\), which depends on the choice of equilibrium quality scores by other proposers. This demand model is described in detail in §3.1. Although for each \(s_q\), the profit margin depends on the specific value of \(b\), the proposer’s demand does not. Therefore, we solve the tagged proposer’s problem in two steps: first we obtain conditional optimal parameter \(b\), and the maximum expected profit margin \(m(s_q, t)\) for each quality score \(s_q\), and then we obtain an equilibrium value of the quality score \(s^*\). Equilibrium quality score is denoted by \(\sigma_q\) and the corresponding total score by \(\rho\).

Both are functions of proposer type \(t\). The tagged proposer chooses its equilibrium quality score by solving the following optimization problem:

\[ \max_{0 \leq s_q \leq s^*} \{ \pi(s_q, t) = m(s_q, t)d(s_q) \} \tag{1} \]

where \(s^*\), the maximum achievable quality score, is the score beyond which every proposer type will earn a negative profit. A proposer will participate only if it makes nonnegative expected profit. This is both a common and a reasonable assumption because proposing entities and service chains are created in response to the BPCI program.

### 3.1. The Demand Model

We focus on an arbitrary proposer, indexed \(i\), in a market with \(n\) proposers. Because CMS adjusts payments according to labor costs in different regions of the country, we use the term market to refer to a combination of a particular DRG and a geographical region.\(^8\) Let \(\sigma_i(t_i), \ldots, \sigma_i(t_{i,n})\) denote equilibrium quality scores of other proposers given type vector \(t_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)\). Within a market, the \(i\)th proposer’s demand, shown in Equation (2), depends on parameters \(\mu_n, \lambda_n\), and \(v_n\). The parameter \(\mu_n\) represents a proposer’s endowment in our model, i.e., the enrollee population it will serve if no proposer offers an improvement in quality; \(\lambda_n\) is the effect of a proposer’s own score on the demand; and \(v_n\) captures the effect of actions by other proposers. We set \(v_n = 0\) if there is only one proposer or beneficiaries are extremely loyal to their providers in a particular market:

\[ d_i(s_q) = E[D_i(s_q, t, \sigma_q, T_{-i})] = [\mu_n + \lambda_n s_q - v_n \sum_{j \neq i} E[\sigma_j(T_j)]G(\rho(T_j))] \tag{2} \]

where \(\rho(T_j) = \gamma \sigma_j(t_j) + b^i(T_j) - b(T_j), \forall j\). In (2), the expectation is over types \(T_j\) of other proposers. Proposer \(j\)’s equilibrium score is multiplied in the last term by \(G(\rho(T_j))\) because this proposer will influence the tagged proposer’s demand only if it is selected. We assume that values of these parameters are such that \(E[D_i(s_q, t, \sigma_q, T_{-i})]\) is nonnegative for \(s_q \in [0, \bar{s}_q]\).

The demand model in (2) assumes that at least some beneficiaries are quality sensitive, which is supported by the following observation in Pearce and Harris (2010, p. VI):

Participating hospitals may also experience increased market share for the service lines … These increases may occur as a result of patient selection of a hospital with perceived higher quality and a more coordinated set of services for their conditions.

Similar observations have also been made in Delisle (2013) and Giles (2011). CMS allows successful proposers to display signage indicating their acceptance into the program, the proposed care pathways, and anticipated quality outcomes for service bundles.

Equation (2) is a descriptive model in which beneficiaries are attracted to networks of providers with higher quality scores and providers compete for beneficiaries via their proposed quality outcomes. Similar descriptive models of price-demand relationship are common in the operations management and marketing literatures; e.g., see (Vohra and Krishnamurthi 2012, p. 41). Higher quality serves the same role as lower price does in those cases. Additional features of our demand model are as follows:

1. Ceteris paribus, a proposer will attract more enrollees when it chooses a higher quality score.
2. We assume that \(\lambda_n \geq (n - 1)v_n\), because if all proposers increase (decrease) their quality scores by one unit, then all proposers will serve more (fewer) enrollees (Tirole 1988).
3. The minimum and maximum demand across all proposers are \(n\mu_n\) and \(n[\mu_n + \lambda_n \bar{s}_q]\), where \(\bar{s}_q\) is the maximum quality score. Maximum demand occurs when the effect of competition is negligible. Note that \(n[\mu_n + \lambda_n \bar{s}_q]\) is not necessarily the entire population of beneficiaries, but rather the
maximum number of enrollees that the particular cohort of proposers can potentially serve. Competition among proposers redistributes enrollees such that
\[ n\mu_n \leq \sum_i E[D_i(s_q, t, \sigma_q, T_n)] \leq n[\mu_n + \lambda_n \sigma_q]. \quad (3) \]

### 3.2. The Payer’s Objective

CMS as its role as a social planner strives to increase the total expected benefit to all beneficiaries. Its objective function can be written as follows:

\[ \psi(G) = \sum_{\text{proposers}} \left( \int \text{(benefit per proposer | type)} P(\text{type}) \right) \]
\[ = \sum_{i=1}^n \int_{q} \rho(t_i) \left[ \mu_n + \lambda_n \sigma_q(t_i) - \nu_n \sum_{j \neq i} E[\sigma_q(T_j)G(p(T_j))] \right] \cdot G(p(t_i)) \, dF(t_i), \quad (4) \]

where \( \rho(T_i) = \gamma \sigma_q(T_i) + b^s(T_i) - b^5(T_i) \) is the equilibrium score, \( \sigma_q(T_i) \) is the equilibrium quality score, and \( b^s(T_i) \) represents the optimal value of parameter \( b \) for type-\( k \) proposers. The payer continues to pay non-BPCI providers in the usual manner. Because there is no change in payments or quality of those providers, they do not affect the payer’s benefit, which is nonzero only when the offered total score is nonzero.

### 3.3. Preliminary Analysis

We first state technical assumptions needed for the ensuing analysis. These assumptions are necessary for tractability. Similar assumptions are made in the vast majority of problems involving the solution of Nash games with incomplete information (see, e.g., Gibbons 1997). In this section, we use \( \phi = g/G \) to denote the reversed hazard rate function of \( G \).

**Assumption 1.** Cost \( c(y, t) \) is twice differentiable and increasing convex in \( y \). Furthermore, the rate at which \( c(y, t) \) increases in \( y \) decreases in \( t \).

**Assumption 2.** There is a monotone ordering of types in the sense that for a fixed quality score, an ordering of types implies an ordering of maximum profit margins.

**Assumption 3.** The function \( G(\cdot) \) is log concave and \( \phi(\cdot) \) is continuous and differentiable.

Mathematically, Assumption 1 means that \( c'(y, t) = \partial c(y, t)/\partial y \geq 0 \), \( c''(y, t) = \partial^2 c(y, t)/\partial y^2 \geq 0 \), and \( \partial^2 c(y, t)/\partial y \partial t \leq 0 \). Assumption 1 is consistent with intuition. Proposers’ costs increase at an increasing rate in quality score, and at each quality level, a higher type proposer experiences a smaller rate of increase in cost as it strives for higher quality. Note that \( c(0, t) > 0 \) for every \( t \), i.e., cost is strictly positive for every proposer type even if it does not propose to improve quality over the historical average. Similarly, Assumption 2 means that the maximum profit margin \( (b^s(t) - c(y, t) - \epsilon_t) \) is increasing in \( t \) for each fixed \( y \). Assumption 3 may be viewed as a regularity condition on \( G \). Many commonly used distributions are log concave; examples include normal, exponential, uniform, and Weibull with shape parameter greater than 1 (see, e.g., Bagnoli and Bergstrom 2005).

The first problem we tackle is the problem of determining \( m(s_q, t) \), the maximum per enrollee expected profit, for a type-\( t \) proposer when it targets a quality score \( s_q \). For a given value of \( s_q \), this problem can be expressed as follows:

\[ m(s_q, t) = \max_{0 \leq b \leq b^s} (b - c(s_q, t) - \epsilon_t)G(\gamma s_q + b^s - b). \quad (5) \]

Define \( s_q^j(t) = \max \{0, [\gamma(\partial c(y, t)/\partial y = \gamma)] \} \) as the highest quality score at which the marginal cost of each unit increase in quality equal \( y \). Let \( b^*(s_q) \) denote the optimal solution for the proposer’s problem for a given \( s_q \). Using arguments from the theory of nonlinear optimization, we can prove the following results. Unless mentioned otherwise, all proofs are provided in the electronic companion.

**Proposition 1.** 1. For each \( s_q \), the corresponding optimal payment \( b^*(s_q) \) is either equal to its upper bound \( b^s \) or a solution of the following first-order equation:

\[ G(\gamma s_q + b - b^*(s_q)) - (b^*(s_q) - c(s_q, t) - \epsilon_t)g(\gamma s_q + b - b^*(s_q)) = 0. \quad (6) \]

2. The optimal payment \( b^*(s_q) \) is a nondecreasing function of \( s_q \). Furthermore, if \( b^s(s_q) \) is a solution of (6), then \( \partial b^*(s_q)/\partial s_q \) is either less than \( \gamma \), or equal to \( \gamma \), or greater than \( \gamma \) depending on whether \( s_q < s_q^j(t) \), or equal to \( s_q^j(t) \), or greater than \( s_q^j(t) \).

3. For each \( s_q \), the maximum expected profit rate equals

\[ m(s_q, t) = \begin{cases} (b^s(t) - \epsilon_t - c(s_q, t))G(\gamma s_q) & \text{if } s_q \geq \hat{s}_q(t) \\ G(\gamma s_q + b^5 - b^*(s_q))^2 & \text{otherwise,} \end{cases} \quad (7) \]

where \( \hat{s}_q(t) \) is either zero or a solution of

\[ \frac{1}{b^s - c(\hat{s}_q(t), t) - \epsilon_t} - \phi(\gamma \hat{s}_q(t)) = 0. \]

Parts 1 and 2 of Proposition 1 establish the intuitively appealing result that a proposer will not pick zero target payments, and that target payments will be nondecreasing in the proposed quality score. Finally, part 3 establishes that there is a maximum quality score, above which the proposer will always choose the highest target payment permitted. There are five immediate corollaries of Proposition 1, which we present next. Corollaries 1 and 2 are presented without a formal proof.
Corollary 1. A type-$t$ proposer will not choose a quality score greater than $\bar{s}_q(t) = c_i^{-1}(b^t - \epsilon_i)$, if it submits a proposal. Furthermore, threshold $\bar{s}_q(t)$ is increasing in $t$.

Note that if the solution of $b^t(t) - c(s_q, t) - \epsilon_i = 0$ is not unique, we select the largest value as $\bar{s}_q(t)$. Thus, type-$t$ proposer must choose $s_q$ such that $0 \leq s_q \leq \bar{s}_q(t)$ because it will earn a negative profit if it chooses $s_q > \bar{s}_q(t)$. We refer to $\bar{s}_q(t)$ as the zero-profit quality score of type-$t$ proposer. Note that $\bar{s}_q$, introduced earlier, is the largest $\bar{s}_q(t)$ across all $t$.

Corollary 2. $\bar{s}_q(t)$ is increasing in $t$.

Corollary 2 is a consequence of Assumptions 1–3. It establishes that the quality threshold beyond which a proposer offers no discount increases with its type. Next, Corollary 3 establishes a technical result concerning $m(s_q, t)$.

Corollary 3. The maximum expected profit margin $m(s_q, t)$ is continuous, everywhere differentiable, and log concave in $s_q$ for each $t$. Furthermore, if $s_q^*(t) \leq \bar{s}_q(t)$, then $\arg\max_{s_q} m(s_q, t) = s_q^*(t)$; otherwise, $\arg\max_{s_q} m(s_q, t) \in (\bar{s}_q(t), s_q^*(t))$.

Corollary 4. $s_q^*(t)$ is increasing in $t$.

Corollary 4 is a consequence of Assumption 1. Costs are increasing in $s_q$ and a higher type proposer’s cost of improving quality increases at a lower rate in quality score. Therefore, the point at which its marginal cost of quality improvement equals $\gamma$ is higher. Next, we prove in Corollary 5 that the tagged proposer’s profit function is quasi-concave in $s_q$.

Corollary 5. The function $\pi(s_q, t)$ is quasi-concave in the target quality score $s_q$ for each $t$.

Corollary 5 follows from Corollary 3 and the fact that the product of log-concave functions is also log concave. The strategy space of a proposer that observes its type $t \in [\bar{t}, \tilde{t}]$ is $s_q \in [0, \bar{s}_q(t)]$. Because the strategy and type spaces of proposers are compact and objective functions are continuous and quasi-concave in their own strategies, a pure strategy Bayesian Nash equilibrium exists for each proposer (Reny 2005).

We next proceed to identify an equilibrium strategy. For this purpose, we assume at first that the tagged proposer, indexed $i$, believes that other proposers will use the function $\sigma_q(\cdot)$ to determine their equilibrium quality scores after observing their types. In Proposition 2, we identify this equilibrium response $\sigma_q(t)$ for the tagged proposer who observes type $t$. In this proposition, and the rest of the paper, we use the prime notation to denote the derivative with respect to $s_q$.

Proposition 2. Under Assumptions 1–3, the following statement is true.

An equilibrium quality score for the tagged proposer, given its type $t$, is either 0 or a solution to the following equation:

$$\lambda_{\gamma} = - \frac{m'(s_q, t)}{m(s_q, t)}.$$  \hspace{1cm} (8)

Because proposers are symmetric, $\sigma_q(t)$ must be a solution of (8) for each $t$.

Corollary 6. For a type-$t$ proposer, if $s_q^*(t) < \bar{s}_q(t)$, then an equilibrium target quality score $\sigma_q(t) \geq s_q^*(t)$. In contrast, if $s_q^*(t) \geq \bar{s}_q(t)$, then the equilibrium quality score $\sigma_q(t) \geq \bar{s}_q(t)$ but it could be either smaller or larger than $s_q^*(t)$.

The intuition behind Corollary 6 is that if $s_q^*(t)$ could be achieved without violating the target payment threshold $b^t(t)$, then the tagged proposer would choose at least $s_q^*(t)$ because that maximizes the profit per enrollee. If not, then the magnitude of the equilibrium quality score cannot be predicted for arbitrary parameter values.

An interesting question arises at this point. Under what conditions would $\sigma_q(t)$ be increasing in $t$? In general, $\sigma_q(t)$ may not be monotone in $t$. However, in the next corollary, we establish a condition for the monotonicity of $\sigma_q(t)$.

Corollary 7. The equilibrium quality score $\sigma_q(t)$ is increasing in $t$ if $b^h(t)$ is increasing in $t$.

Corollary 7 may lead one to think that discount $(b^h(t) - b^*(t))$ may be increasing in type (profit margin) when $\sigma_q(t)$ is increasing in $t$. That is, however, not true because a proposer’s equilibrium quality score in type $t$ may switch between either being greater or less than $\bar{s}_q(t)$ several times as $t$ increases. For example, suppose $t_1 \geq t_2$ and $s_q^*(t_1) \geq \bar{s}_q(t_1) \geq \bar{s}_q(t_2) \geq s_q^*(t_2)$. In that case, we can have $\sigma_q(t_1) \geq \sigma_q(t_2)$ if $b^h(t_1) \geq b^h(t_2)$. However, we have $b^h(t_1) - b^*(t_1) = 0 \leq b^h(t_2) - b^*(t_2)$, i.e., a proposer with higher profit margin may offer a lower discount.

Given $\sigma_q(t)$ as an equilibrium strategy, the payer’s objective function in (4) can be rewritten as

$$\psi(G) = \sum_{i,j} \rho(t_i) \left[ \mu_{\gamma} + \lambda_{\gamma} \sigma_q(t_i) - \nu_{\gamma} \sum_{j \neq i} \sigma_q(t_i) \right] G(\rho(t_i))dF(t_i) + G(\rho(t_i))dF(t_i).$$  \hspace{1cm} (9)

where $\rho(t_i) = \gamma \sigma_q(t_i) + b^h - b^*(\sigma_q(t_i))$ $\forall i$. The payer can affect proposers’ rational expectation about $G$ by selectively revealing some information. The payer’s problem is complicated because of two reasons.

First, it is not clear whether proposers will increase or decrease equilibrium quality scores in response to different expectations about the probability of being selected. Second, even if proposers’ responses are monotone, e.g., increasing in payer’s actions leading to different levels of...
revealed information, it is unclear if the payer will benefit from such actions. On the one hand, fewer proposers may be selected, which may result in fewer total number of enrollees served. On the other hand, those proposers that are selected will have higher quality scores, increasing benefit per enrollee served.

4. BPCI Mechanism: Analysis and Insights

In this section our goal is to ascertain whether payers such as CMS should eliminate selection uncertainty by announcing a fixed threshold. Our main result in §4.1 is that the payer can benefit from uncertain selection and this happens because of either information asymmetry or competition. In §4.2, we determine how a payer should choose the fixed threshold for those problem scenarios in which a fixed threshold would be desirable.

4.1. Uncertain or Fixed Threshold Mechanism?

The problem of deciding which mechanism to use is non-trivial because of two complicating factors. First, one needs to identify optimal uncertain and optimal fixed threshold mechanisms. The former is particularly difficult because one may not restrict attention to a particular class of probability distributions. Arbitrary distributions may have an infinite number of parameters, making the problem immediately intractable. The second difficulty is that even within a particular class of distributions, it is not apparent how proposers’ equilibrium responses would change if some parameters of the distributions were changed. Presenting details of our analysis, we provide a high-level description of our approach to address these difficulties.

We assume that the payer chooses an arbitrary distribution \( G(\cdot) \) by revealing some information to proposers, and focus on two special cases. In one instance, to isolate the effect of information asymmetry, we consider a single proposer whose type could be one of a continuum of types. This instance corresponds to markets in which either there is only one entity that is capable of convening multiple providers to offer bundled services for a particular DRG, or beneficiaries are extremely loyal to their providers, or both. In either case, each proposer’s decision problem can be analyzed independently, resulting in the single proposer model. The second instance has multiple proposers competing with each other for attracting beneficiaries. However, all proposers are of the same type and have no information asymmetry. This corresponds to a homogenous market of proposers.

For the single-proposer model, we first identify properties of \( G(\cdot; \theta) \), parameterized by \( \theta \), such that \( r^*(\theta) \) can be predicted where \( r^*(\theta) \) is the optimal score of the proposer. In particular, we identify conditions under which the proposer’s response is monotone in \( \theta \). Then, we show that there exist cases in which \( \psi(G(r^*(\theta^*)); \theta^*) > \psi(r^*) \). Recall that the latter is the maximum benefit that the payer can realize by choosing an optimal fixed threshold \( r^* \). Because \( G \) represents an arbitrary distribution, which is not necessarily parameterized by \( \theta \), the above inequality immediately allows us to conclude that \( \psi(G^*) \geq \psi(G(r^*(\theta^*))); \theta^* > \psi(r^*) \), where \( G^* \) denotes the optimal distribution. An optimal selection mechanism in such cases cannot be a fixed threshold.

For the single-proposer-type model, we consider two simple forms of \( G \)—either the probability of acceptance is a constant (independent of the proposed score), or linearly increasing in the proposed score. Our analytical results pertain to the former. We identify conditions under which the optimal constant acceptance probability, denoted by \( \eta^* \), is such that \( \psi(\eta^*) > \psi(r^*) \), which immediately implies that \( \psi(G^*) \geq \psi(\eta^*) > \psi(r^*) \). Similar arguments can be given when acceptance probabilities are linearly increasing in score.

4.1.1. Single Proposer. For the model presented in this section, proposer type \( t \in [t_1, t_\bar{1}] \), and \( n = 1 \) and \( \nu_1 = 0 \) in the demand function \( d(s) \) given in (2).

The payer’s objective function in (9) simplifies significantly when there is only one proposer as shown below:

\[
\psi(G) = \int_{t_1}^{t_{\bar{1}}} r^*(t)(\mu_1 + \lambda_1 s^*_t(t))G(r^*(t); \theta)dF(t). \tag{10}
\]

Note that we use \( r^*(t) \) and \( s^*_t(t) \) to denote the proposer’s optimal (not equilibrium) responses because it faces no competition. Suppose \( G(\cdot) \) is characterized by a parameter \( \theta \), which could be affected by the payer’s actions. Furthermore, suppose \( \phi(r; \theta) = g(r; \theta)/G(r; \theta) \) is increasing in \( \theta \), which implies an increasing reversed hazard rate ordering\(^{11} \) of \( G \) in \( \theta \). Among distributional forms that may be used for \( G(\cdot) \), possible choices include exponential, Gamma, Weibull, log normal, and uniform because each of these distributions is both log concave and exhibits reverse-hazard rate ordering in terms of one or more of its parameters—see Table 1.1 of Müller and Stoyan (2002) for details.

**Proposition 3.** If the payer’s actions result in increasing (respectively, decreasing) reversed hazard rate order of the probability distributions of acceptance, then proposers’ optimal quality scores \( s^*(t) \) change as follows:

1. if \( s^*_t(t) < s^*_\bar{t}(t) \), then \( s^*(t) \) decreases (respectively, increases) in \( \theta \); else
2. if \( s_\bar{t}(t) \leq s^*_t(t) \leq s_\bar{t}(t) \), then \( s^*(t) \) increases (respectively, decreases) in \( \theta \).

An immediate corollary of Proposition 3 is that \( r^*(t) \) is monotone in \( \theta \).

**Corollary 8.** If the payer’s actions result in increasing (respectively, decreasing) reversed hazard rate order of the probability distributions of acceptance, then \( r^*(t) \) is increasing (respectively, decreasing) in \( \theta \).
To understand the above results on an intuitive level, consider the case when $\phi$ is increasing in $\theta$. This means that if the proposer believes that it will succeed upon offering a total score of $r$ or less, its instantaneous chance of winning if it offers precisely $r$ is increasing in $\theta$. This incentivizes proposers to select higher total scores as $\theta$ increases, which explains Corollary 8. To understand Proposition 3 on an intuitive level, recall from the definition of $\hat{s}_1(t)$ in Proposition 1 that if $s'(t) > \hat{s}_1(t)$, then $b'(t) = b^*(t)$ must be true. Therefore, a proposer who wants to increase total score in response to an increase in $\theta$ has only one option— increase quality score. In contrast, when $s'(t) < \hat{s}_1(t)$, we have from Proposition 1 that $b'(t) < b^*(t)$ must hold. In this case, if the proposer increases $s'(t)$, then for each unit increase, if it wants to maintain the same profit per enrollee, it must concomitantly increase $b^*(t)$ by an amount greater than $\gamma$. Following these actions, total score will not increase. Therefore, a better strategy is to decrease $s'(t)$ and also $b'(t)$.

Suppose that different levels of information sharing indeed produce monotonically reversed hazard rate order of the probability distributions of acceptance. Then, we can write down an equation whose solution determines an optimal information sharing level. We develop the necessary arguments below for situations in which information sharing levels induce an increasing reversed hazard rate order on $G$. Similar arguments can be developed if the opposite is true. It is well known that the reversed hazard rate order implies the usual stochastic order. That is, if $\phi(r; \theta_1) \leq \phi(r; \theta_2)$ for every $r$ when $\theta_1 \leq \theta_2$, then $G(r; \theta_1) \geq G(r; \theta_2)$ for every $r$ ($G$ decreases in $\theta$). Next, taking the derivative of the integrand in (10), where the integrand is denoted by $\Psi(t; \theta)$, we obtain

$$
\frac{\partial \Psi(t; \theta)}{\partial \theta} = \left( (\mu_1 + \lambda_1 s'(t)) G(r^*(t); \theta) \right) \frac{\partial r^*(t)}{\partial \theta} + r^*(t) \frac{\partial}{\partial \theta} \left( (\mu_1 + \lambda_1 s'(t)) G(r^*(t); \theta) \right).$$

(11)

When $\phi(r; \theta) = g(r; \theta)/G(r; \theta)$ is increasing in $\theta$, $\partial r^*(t)/\partial \theta \geq 0$, $\partial G(r; \theta)/\partial \theta |_{r=r^*(t)} \leq 0$ and the sign of $s'(t)/\partial \theta$ depends on whether $s'(t) < \hat{s}_1(t)$ or not. Hence, the first term in (11) is always positive, whereas the second term can be positive or negative, which suggests that either an extreme value of $\theta$ will be optimal or an optimal value can be obtained by equating the above partial derivative to zero for all types. On an intuitive level, suppose $\phi$ is monotone in $\theta$. Then, information that causes $G$ to increase (decrease) in reversed hazard rate order will cause each proposer type to offer a higher (lower) total score, but the expected number of proposers selected (and hence, the expected number of beneficiaries served) may be lower (higher). The net effect of these two opposing forces is not necessarily such that the payer should eliminate selection uncertainty. We present an example to illustrate the optimality of the uncertain mechanism next.

Suppose the proposer may belong to one of two groups determined by $t \in [\tilde{t}, \bar{t}]$ or $t \in [\bar{t}, \tilde{t}]$ such that all proposer parameters within a group are identical. Furthermore, let $F(\bar{t}) = 0.9$, $\mu_q = 1$, $\lambda_1 = 3$, $\gamma = 0.6$, and the zero-profit quality scores of the two groups of proposer types be $\tilde{s}_1 = 5$, $\tilde{s}_2 = 17$. If the payer were to choose a fixed threshold, its optimal strategy would be to set $r^* = 5\gamma$. The proposer would then respond either with a score $s_1^* = 5$ (if $t \in [\tilde{t}, \bar{t}]$), or with a score $s_2^* = 8.33$ (if $t > \bar{t}$). The payer’s benefit would be 56.2. In this base case, $G_1(r) = 0$ if $r < 5\gamma$, and $G_2(r) = 1$ otherwise. In contrast, if its information sharing efforts were to have the proposer believing that its chances of success would be $G_2(r) = 0$ if $r < 5\gamma$, $G_2(r) = 0.9754$ if $5\gamma \leq r < 9\gamma$, and $G_2(r) = 1$ otherwise, then the proposer upon drawing $t \in [\tilde{t}, \bar{t}]$ would respond with a score of $s_1^* = 5$, and a score of $s_2^* = 9$, otherwise. The payer’s benefit would be 57.26. That is, the payer would benefit from having an uncertain selection mechanism. It is easy to verify that $G_2(r)/G_1(r)$ is increasing in $r$ in this example in the range of $r$ values for which $G_1(\cdot)$ are nonzero.

Before closing this section, we wish to emphasize that the proposer faced no competition under the assumptions made in this section. Still, the payer might find it optimal to use an uncertain mechanism. This result is a consequence of information asymmetry. In what follows, we analyze special cases in which proposers have discrete types and linear costs. By constructing a different model with a single type of proposers, i.e., with no information asymmetry, we show in the sequel that competition for beneficiaries is another reason why a payer may prefer an uncertain selection mechanism.

4.1.2. Single-Proposer Type. Suppose we have $n$ potential proposers of a single type (we omit subscript $i$ in our subsequent discussion). Costs are linear, i.e., $c(s) = \alpha s + \beta$, where $\alpha \geq 0$, $\beta > 0$. We use $\eta$ instead of $G(\cdot)$ to denote the probability of being selected. We consider two possible forms of $\eta$. In one case, $\eta$ is a constant that does not depend on a proposer’s score. In the second case, $\eta$ is linearly increasing in offered score. Before proposers submit proposals, the payer releases information that impacts the beliefs of proposers about $\eta$, but there is no fixed threshold that guarantees success.

Consider a tagged proposer and omit subscript $i$ in the preliminary analysis. The tagged proposer’s expected profit margin maximization problem (when $\eta$ is a constant) in Equation (5) simplifies as follows:

$$m(s_q) = \max_{0 \leq b \leq b^*} [b - \alpha s_q - \beta - \epsilon] \eta.$$  

(12)

For a fixed $s_q$, $m(s_q)$ is increasing in $b$. Hence, we have $b^* = b^h$ and $m(s_q) = (b^h - \alpha s_q - \beta - \epsilon) \eta$. Also, profit rate is nonnegative only when the selected quality score is $s_q \leq (b^h - \beta - \epsilon)/\alpha$.

The above arguments imply that the zero-profit quality score for the tagged proposer, $\tilde{s}_q = (b^h - \beta - \epsilon)/\alpha$. Thus,
the maximum score that can be achieved by the tagged proposer is \( \bar{r} = \bar{y} \bar{s}_q \).

Because proposers have no private information, the payer could require proposers to achieve a particular score. However, because we focus in this section on the BPCI mechanism, the proposal submission and selection processes are not altered (we discuss an optimal contract in \( \S 5 \)). That is, each proposer selects its quality score. Proposer \( i \)'s demand when it chooses quality score \( s_q \) and other proposers choose an equilibrium score \( \sigma_q \), is \( d_i(s_q, \sigma_q) \). If the payer decides to use a fixed threshold with single type, it will choose \( \bar{r}^{\text{fu}} = \bar{r} \) because in that case all \( n \) proposers will participate and each proposer will offer its maximum score. Alternatively, the payer may choose an uncertain selection mechanism. Selection probabilities either may be independent of score, or depend on offered score according to some function. We provide details for the case that has a constant score, or depend on offered score according to some function.

We provide details for the case that has a constant score, or depend on offered score according to some function. We can calculate equilibrium score and expected demand of selection. Moreover, the uncertain selection mechanism is relatively small, and (2) the effect of a proposer’s own quality score on its demand is moderate. These statements follow from the bounds on the values of \( \mu_n, \lambda_n \) and \( \nu_n(n-1) \) in the proposition.

Because we chose an arbitrary uncertain mechanism and found it to be better for the payer relative to an optimal fixed threshold, this implies that the optimal uncertain mechanism must be better as well. We illustrate the impact of uncertain selection with the help of the following example. In the example, \( n = 2 \), \( \nu_n = \mu_n = 1 \), \( \gamma = 0.6 \), \( \bar{s}_q = 15 \), and \( \lambda_n = 1.05 \). For the parameters in this example, we have \( (n-1)\nu_n = 1 < \lambda_n = 1.05 < 2(2 - \sqrt{2})(n-1)\nu_n = 1.17 \) and \( \mu_n = 1 < 3\lambda_n + 2(2 - \sqrt{2})(n-1)\nu_n = 1.82 \). Therefore, per Proposition 4, uncertain mechanism is better for the payer. The optimal value of the payer’s objective function \( \psi \) is 42.08 (respectively, 31.5) if uncertain selection mechanism (respectively, fixed threshold) is utilized. The optimal value of \( \eta \) is 0.761. The optimal fixed threshold equals \( \bar{s}_q = 15 \) and proposer quality scores equal \( \bar{\sigma}_q = 15 \). Under uncertain selection, proposers quality scores \( \bar{\sigma}_q = 11.01 < \bar{s}_q \). Still the uncertain mechanism is better because the expected demand per proposer is greater if that mechanism is used (3.185 versus 1.75 with fixed threshold).

The payer may release information that selection probability will increase with the proposed score (effectively, proposer types). For example, \( \eta \) may linearly increase in the proposed score with \( \eta = \rho / (\bar{r} - r) \), where \( \rho = \gamma \bar{\sigma}_q \). We can calculate equilibrium score and expected demand in that case as well, although the analysis does not result in closed-form expressions similar to (14)–(16). Upon reworking the example above with \( r = 0 \) and \( \eta = \rho / \bar{r} \), we find that payer’s benefit improves relative to the constant probability of selection. Moreover, the uncertain selection mechanism remains better than the fixed threshold (examples can be found in \( \S 5 \)).

The key finding in this section is that in the presence of competition for beneficiaries, even when proposers know each other’s type, there are situations in which the payer should not announce a fixed threshold. This can be explained as follows. Our demand model endows each potential proposer with a baseline demand from beneficiaries. The total number of beneficiaries is much larger than the sum of the endowments of all potential proposers. Proposers gain beneficiaries by poaching from other proposers as well as by attracting beneficiaries who are not included in any proposer’s endowment. There could exist market conditions in which each proposer has a relatively small endowment and beneficiaries’ choices are more sensitive to proposed scores when there are fewer bundled providers—i.e., exclusivity matters. The uncertain mechanism reduces the expected number of proposers that are eventually selected. Although the selected proposers have lower scores relative to the case when all proposers are...
4.2. Multiple Proposers and Types

Suppose as in §3, we have \( n \) potential proposers, each of which can be one of \( \ell \geq 1 \) discrete types, i.e., \( t \in \{1, \ldots, \ell\} \). Costs are linear, i.e., \( c(s, t) = c_\ast + \beta_t \), where \( \alpha > 0, \beta > 0 \), and consistent with Assumption 2, the maximum margin \( b^*(t) = c(s, t) - \epsilon \) is increasing in \( t \). Each proposer’s type is \( t \) with probability \( p_t \), independently of other proposers’ types. Note that we use \( \{p_t\} \) rather than \( F(\cdot) \) in this section to highlight the fact that proposer types are discrete.

Assumptions 1–2 of §3 are applicable to this model. The optimal value of \( \kappa^\ast \), given \( s_j \), can be obtained in a manner similar to §4.1.2, with the difference that we add subscript or argument \( t \) when referring to a type-\( t \) proposer. Thus, we have \( b^*(t) = b^\ast(t) \) for every \( s_j \) and the zero-profit quality score \( \bar{s}_j(t) = (b^\ast(t) - \beta_t - \epsilon_t) / \alpha \) and zero-profit total score \( \bar{r}(t) = \gamma \bar{s}_j(t) \). It should be clear that because profit margin is increasing in \( t \), \( \bar{s}_j(t) \) and \( \bar{r}(t) \) are also increasing in \( t \).

Using the framework presented in Proposition 2, a proposer can determine its equilibrium quality score \( \sigma_q(t) \) for each type \( t \). Notation \( r^a \) in this section denotes the payer-announced fixed threshold.

**Proposition 5.** Suppose \( r_{j-1} < r^a \leq \bar{r}_t \) and let \( t (j \leq t \leq \ell) \) be the largest type such that,

\[
\left( (\bar{s}_j(t) \lambda_n - \mu_n) \right) \left( 2\lambda_n - \nu_n(n-1) \sum_{k=t+1}^\ell p_k \right) + (n-1)\nu_n \sum_{k=t+1}^\ell (\bar{s}_j(k) \lambda_n - \mu_n) p_k \cdot \left( 2\lambda_n (2\lambda_n - \nu_n(n-1) \sum_{k=j}^\ell p_k ) \right)^{-1} \leq \frac{r^a}{\gamma} \leq \bar{s}_j(j),
\]

then an equilibrium score \( \sigma_q(t) \) can be characterized as follows:

\[
\sigma_q(v) = 0, \quad v = 1, \ldots, j-1,
\]

\[
\sigma_q(v) = \frac{r^a}{\gamma}, \quad v = j, \ldots, t, \text{ and }
\]

\[
\sigma_q(v) = \frac{1}{2\lambda_n (2\lambda_n - \nu_n(n-1) \sum_{k=t+1}^\ell p_k )} \cdot \left( (\bar{s}_q(v) \lambda_n - \mu_n) \right) \left( 2\lambda_n - \nu_n(n-1) \sum_{k=t+1, k \neq v}^\ell p_k \right) + (n-1)\nu_n \sum_{k=t+1, k \neq v}^\ell (\bar{s}_q(k) \lambda_n - \mu_n) p_k + 2\lambda_n \frac{r^a}{\gamma} \sum_{k=t}^\ell p_k \right) \cdot \left( 2\lambda_n (2\lambda_n - \nu_n(n-1) \sum_{k=j}^\ell p_k ) \right)^{-1} \leq \frac{r^a}{\gamma} \leq \bar{s}_j(j),
\]

where \( \sigma_q(t) \) depends on \( r^a \). Our next result concerns the optimal value of the threshold score for the payer.
PROP. 6. The optimal threshold \( r^* \in \{ \tilde{r}_k, k = 1, 2, \ldots, \ell \} \). That is, the payer’s objective function is maximized by selecting one of the zero-profit scores of the \( \ell \) proposer types if it chooses to announce a fixed threshold.

Intuitively, the result in the above proposition will be trivial to argue if between any two consecutive zero-profit scores, the payer’s benefit from group 2 and group 3 proposers is either simultaneously increasing or simultaneously decreasing in \( r^* \). Corollaries EC.1–EC.2 in the electronic companion establish that this property may not hold; in particular, payer’s benefit from group 2 proposers increases but that from group 3 proposers may decrease. To prove Proposition 6, we therefore show that the rate of increase of payer’s benefit from group 2 proposers is greater than the rate of decrease from group 3 proposers. Thus, rather than choose \( \tilde{r}_{j-1} < r^* \leq \tilde{r}_j \), a payer will prefer to select \( r^* = \tilde{r}_j \). However, the payer’s objective function is discontinuous as \( r^* \) crosses each \( \tilde{r}_j \) because some proposers no longer participate. Therefore, which zero-profit score will be optimal requires a search over a finite number of choices.

5. Constrained Optimal Mechanism

For our analysis in this section, we assume that proposer types are discrete, their costs are linear (as in §§4.1.2 and 4.2), and the relative quality weight \( \gamma \) is fixed. Clearly, the optimal contract with a fixed \( \gamma \) is not an overall optimal mechanism. Because CMS is a social planner in our model, we do not treat \( \gamma \) as a decision variable. Therefore, our analysis can be viewed as a search for a constrained optimal mechanism.

We consider contracts involving three parameters: quality score \( s \), bundled payment \( b \), and probability of success \( \eta \). Following the revelation principle, the payer may restrict attention to a direct mechanism involving at most \( \ell \) contracts—one for each proposer type (Myerson 1981). Consider the following direct mechanism. The payer offers a menu of contracts to potential proposers. Proposers submit their types to the payer who then allocates contracts based on the revealed types. Contracts are designed such that each proposer maximizes its profit by revealing its true type. How should the payer select contract parameters? We address this question in the sequel.

Let \( \mathcal{P} \) be the set of all proposer types. The payer maximizes its benefit from each proposer by solving the following optimization problem:

\[
\psi = \max_{(s_i, b, \eta_i)_{i \in \mathcal{P}}} \sum_{i \in \mathcal{P}} p_i(\gamma s_i + b^h(i) - b) d(s_i) \eta_i
\]

subject to

\[
0 \leq s_i \leq S_n(i), \quad 0 \leq b_i \leq b^h(i), \quad 0 \leq \eta_i \leq 1, \quad \forall i \in \mathcal{P}
\]

\[
\pi(s_i, b_i, \eta_i|T=i) \geq 0, \quad \forall i \in \mathcal{P}
\]

\[
\pi(s_i, b_i, \eta_i|T=i) \geq \pi(s_j, b_j, \eta_j|T=i), \quad \forall j \neq i
\]

and \( \forall i, j \in \mathcal{P} \),

where \( \pi(s_i, b_i, \eta_i|T=i) \) is the expected profit of a proposer that picks contract \((s_i, b_i, \eta_i)\) upon learning that its type is \( i \), \( \forall i \) and \( j \), and \( d(s_i) = \mu s_i - \nu s_i(n-1) \sum_{j=1}^{\ell} p_j s_j \eta_j \) is the expected demand of a type-\( i \) proposer when it picks contract menu \( i \). Note that the above formulation allows \( \gamma s_i + b^h(i) - b \cdot \eta_i \) to equal zero. When that happens, type-\( i \) proposers do not participate. Note that we use \( i \) as type index instead of proposer index for our discussion in this section.

Constraints (23) specify relevant bounds on the values of contract parameters, whereas Constraints (24) and (25) are known as individual rationality (IR) and incentive compatibility (IC) constraints, respectively. The IC constraints are not required when the payer knows each proposer’s type, i.e., proposers do not possess private information. The corresponding solution is an important benchmark known as the first-best solution. We analyze this case in Appendix B. Our key results (presented in Proposition 7) are that the quality scores are increasing in proposer type, and that in general, the first-best option may involve selection uncertainty. Furthermore, it is optimal to use a fixed threshold when either there is no competition among proposers \( (\nu_s = 0) \), or when quality weight is higher than cost per unit of quality improvement \( (\gamma \geq \alpha) \), each proposer has a relatively large demand endowment \( \mu_s \), and the effect of the proposer’s own quality score on its demand \( \lambda_s \) is relatively large.

When proposers possess private information, constraints (25) cannot be ignored. To obtain an analytical solution in this instance, we make additional monotonicity assumptions; namely, (1) \( b_i + \varepsilon_s \geq b_{i+1} + \varepsilon_{i+1} \), and (2) \( b^h_i \geq b^h_{i+1} \). Our key results (Propositions 8–9), presented in Appendix C, specify properties of optimal contracts. In particular, similar to the first-best solution, the optimal mechanism may, in general, involve selection uncertainty. Moreover, the conditions under which a fixed threshold is optimal are similar to those for the first-best solution.

In what follows, we consider three examples with two proposers and two types and parameters described below.
The optimal mechanism performs better relative to the BPCI mechanism. Why does the optimal mechanism perform better relative to the BPCI mechanism? Example 1 reveals possible reasons. The BPCI mechanism may choose either a type-independent (constant) or type-dependent (linear) selection probability or a fixed threshold. However, it does not simultaneously require proposers to achieve a certain score. In contrast, the optimal mechanism has type-dependent selection probability, a prescribed quality score, and a prescribed discount, which are generally greater than what proposers choose voluntarily.

6. Concluding Remarks and Policy Implications

CMS’ bundled payments program creates incentives for better coordination of services. Successful proposers benefit from the relaxation of laws governing gain sharing. A key challenge for government agencies is determining which proposers should be allowed to participate in such programs and whether they should establish firm selection standards. CMS has chosen an uncertain selection mechanism. In this paper, we analyze whether selection uncertainty may be optimal from a normative modeling perspective.

Our analysis in §§3 and 4 shows that higher profit margin proposers may select higher quality scores if their historical average payments are also higher, but they do not necessarily offer higher discounts. Up to a certain cutoff value of the quality score, which is increasing in proposer type, proposers offer quality improvements and discounts over and above the mandatory discount. Beyond this cutoff, proposers offer only quality improvements and the mandatory discount. However, there may not exist a type threshold such that proposers above or below that threshold offer discounts and others do not. Furthermore, if the costs of proposers are linear, the payer should expect quality improvements and only the mandatory discount.

One of our key findings is that an uncertain mechanism can outperform the fixed selection mechanism. This favors the use of the uncertain selection mechanism. We explain why this happens with the help of two special cases. In one case (§4.1.1), we consider a single proposer with multiple types and in the other (§4.1.2), we consider multiple proposers with a single type. In the first case, there is no competition and in the second, there is no private information (i.e., proposers know each other’s type). We find that uncertain mechanism can be better than announcing a fixed threshold in each case. However, the reasons are quite different—information asymmetry in the first case and the effect of exclusivity on the total number of beneficiaries attracted to all BPCI providers in the second. Note that selection uncertainty is endogenous in our model. It is not caused by CMS’ inability to assess proposal quality.

In addition to explaining why payers may choose an uncertain selection mechanism, we also analyze the situation in which a payer may prefer to announce a fixed threshold. For such cases, we show that when proposer types are discrete and costs are linear, an optimal fixed threshold must be one of the zero-profit scores associated with a proposer type. This result is satisfying on an intuitive level and makes it easy to implement a fixed threshold strategy.

Another key finding of this paper is that the BPCI mechanism is not an optimal mechanism, even when we restrict

---

**Table 1.** Optimal vs. the BPCI mechanism.

<table>
<thead>
<tr>
<th>Example</th>
<th>Constant probability</th>
<th>Linear probability</th>
<th>Fixed threshold</th>
<th>Optimal mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta^<em>$ ($\sigma_1, b_{c_1}^</em>$)</td>
<td>$\eta^<em>$ ($\sigma_1, b_{c_2}^</em>$)</td>
<td>$\eta^<em>$ ($\sigma_1, b_{c_3}^</em>$)</td>
<td>$\eta^<em>$ ($\sigma_1, b_{c_4}^</em>$)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.38</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.37, 0.2)</td>
<td>(0.63, 0.25)</td>
<td>(1, 0.25)</td>
<td>(1, 0.25)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.88</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.4, 0.2)</td>
<td>(0.64, 0.25)</td>
<td>(2, 0.25)</td>
<td>(0.88, 0.19)</td>
</tr>
<tr>
<td>3</td>
<td>0.76</td>
<td>43.23</td>
<td>10.2</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>(11.06, 2, 12.06)</td>
<td>(12.20, 2, 13.43)</td>
<td>(17.21, 2)</td>
<td>(15.2, 16.58)</td>
</tr>
</tbody>
</table>

---

**Example 1.** $n = 2$, $\ell = 2$, $\gamma = 0.6$, $\delta_{2}(2) = 2$, $\delta_{1}(2) = 1$, $\lambda_{1} = 3$, $\mu_{1} = 1$, $\nu_{n} = 0.5$, $p_{1} = 0.8$, and $p_{2} = 0.2$. Cost parameters are selected such that $\alpha = 0.1$, $\beta_{1} = 0.1$, $\beta_{2} = 0.05$, $b^{h}(1) = 0.2$, and $b^{h}(2) = 0.25$.

**Example 2.** $p_{1} = 0.25$ and $p_{2} = 0.75$, and all other parameters are identical to Example 1.

**Example 3.** $n = 2$, $\ell = 2$, $\gamma = 0.6$, $\delta_{2}(2) = 17$, $\delta_{1}(2) = 15$, $\lambda_{1} = 1.05$, $\mu_{1} = 1$, $\nu_{n} = 1$, $\alpha = 0.1$, $\beta_{1} = 0.5$, $\beta_{2} = 0.4$, $b^{h}(1) = 2$, $b^{h}(2) = 2.1$, $p_{1} = 0.9$, and $p_{2} = 0.1$.
attention to the two key attributes of quality and cost and their weights currently used in scoring proposals. We show in §5 that for scenarios with discrete types and linear costs, a constrained optimal mechanism has a tiered structure with three elements—a minimum score, a maximum per episode payment, and a selection probability. The BPCI mechanism, in contrast, does not explicitly link selection probabilities with prescribed minimum scores. We find that prescribed scores under an optimal mechanism are generally greater than those that proposers will choose voluntarily. We also identify conditions under which it will be optimal to announce a fixed threshold. This happens when either proposers do not compete for beneficiaries, or when quality weight is higher than cost per unit of quality improvement, each proposer has a relatively large demand endowment, and the effect of the proposer’s own quality score on its demand is relatively large.

A practical appeal of the uncertain selection mechanism is that it does not require the payer to know the distribution of proposer types; hence, it may offer a hedge between eliminating too many proposer types from the pool relative to a high fixed threshold, and lowering the proposed scores too much relative to a low fixed threshold. In addition to potentially greater societal benefit, an uncertain mechanism may have additional benefits discussed in auction literature, e.g., discouraging collusion, and increasing flexibility to ex post adjust supply in different markets. Its disadvantage is that proposers may question fairness and unbiasedness of expert panels.

This paper has its limitations. First, we model a particular mechanism used by CMS and the results we obtain may not generalize to other procurement mechanisms. Second, we assume that given a quality target, proposers can identify an optimal (minimum cost) service chain design. This is a difficult problem, which may require several attempts (trial and error) in practice until proposers can identify an optimal design. Third, our approach utilizes a descriptive demand model. The demand model is not based on empirical evidence. It captures the expected behavior of beneficiaries, given differences in quality of bundles offered by different proposers. Fourth, our goal in the paper is to determine whether in a normative setting, the payer should select an uncertain mechanism. It is not to comment on whether the mechanism used by CMS performs better than a deterministic mechanism in practice.

Some of the limitations identified above provide opportunities for future work, much of which will depend on the availability of data from CMS and service providers. For example, research could be undertaken to identify approaches that lead to minimum-cost designs for target quality levels. These approaches may differ from one DRG to another. When data on beneficiaries’ choices become available, future efforts may characterize the actual demand function observed in practice. It will be interesting to find out what factors determine the extent to which beneficiaries stick to their “usual” service providers. In a similar vein, BPCI program data may be used to build econometric models to test whether uncertain mechanism actually performed better than the deterministic approaches used by CMS in other similar attempts. We expect the last task to be difficult because in previous attempts the DRGs and bookends were prescribed by the CMS, and gain sharing between hospitals and physicians was not permitted. Finally, there is speculation that bundles (and possibly service chain designs) may be standardized after completion of BPCI contract terms and payments will be limited to bundled prices. Also, there may be some ratcheting effect on prices over time. Modeling such dynamic effects of implementing bundled payment schemes is a topic for future research.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2015.1403.

Acknowledgments

The authors extend thanks to the area editor, the associate editor, and referees for their comments that have helped improve the paper. The authors are also grateful to Milind Dawande, Kingshuk Sinha, and Eli Snir for providing helpful feedback. The authors thank seminar participants at the Johns Hopkins University, the University of Florida, and the University of Illinois at Urbana–Champaign for their comments on earlier versions of the manuscript. This material is based upon work completed, in part, while the first author was on detail at the National Science Foundation under the terms of an Intergovernmental Personnel Assignment [Award CMMI-1452877]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. During the execution of this project, the second author was supported, in part, by a research grant provided by the Office of the Vice President for Research, University of Minnesota [Proposal 22772].

Appendix A. Institutional Background

Models I through III are called retrospective payment models and Model IV is called the prospective payment model (see Center for Medicare and Medicaid Innovation 2011, Scamperle 2013). A common feature of all models is that the target payment must include a mandatory minimum discount to CMS relative to historical average payments for services included in the bundle. In Models I through III, all entities comprising the proposer (e.g., hospitals and physicians) continue to bill for their services on a fee-for-service basis. An audit at the end of the contract period determines the actual average amount paid per episode relative to the target. A fraction of the savings relative to the target are then given to the proposer to be shared among its constituent groups (Berman 2013). If the per-episode payment exceeds the target amount, then the proposer must pay CMS. The three models under this scheme differ in terms of which services may be linked. The possibilities are acute hospital stay only (Model I), acute hospital stay plus post-acute care (Model II), and postacute care only (Model III). In contrast, Model-IV concerns acute hospital stay only and proposers receive a single payment equal to their proposed targets.
Appendix B. The First Best Solution

Our goal in this subsection is to find optimal contracts \((s_i^*, b_i^*, \eta_i^*)\) given \(\beta_i\). The 4th contract is offered to each type-\(i\) proposer on a take-it-or-leave-it basis. We summarize our result in Proposition 7.

Proposition 7. The following statements are true.

1. The IR constraint is tight for all proposer types, i.e., \(b_i^* = \alpha s_i^* + \bar{\beta}_i + \bar{\epsilon}_i\), and every proposer makes zero profit.
2. If for some \(i\), \(s_i^* = 0\), then \(\eta_i^* = 1\) and \(b_i^* = \beta_i + \epsilon_i \leq b^h(i)\). That is, proposers that are asked to offer maximum discount are selected for sure.
3. If for some \(i\), \(0 < s_i^* < \bar{s}_i(i)\), then \(b_i^* < b^h(i)\) and \(\eta_i^* \in [0, 1]\).
4. If for some \(i\), \(s_i^* = \bar{s}_i(i)\), then \(b_i^* = b^h(i)\) and \(\eta_i^* \in [0, 1]\).
5. \(s_i^*\) is increasing in \(i\).
6. If \(\nu = 0\), i.e., proposers’ demands are independent, then every proposer type is selected with probability 1.
7. If \(\gamma > \alpha\), \(\mu \geq 2\bar{s}_i(\ell)\nu_s(n - 1)\) and \(\lambda \geq \nu_s(n - 1) - \bar{s}_i(\ell)/\bar{s}_i(1)\), then every proposer type is selected with probability 1, and \(s_i^* = \bar{s}_i(i)\) and \(b_i^* = b^h(i)\).

Appendix C. Private Information

Proposition 8. Offered contracts \((s_i^*, b_i^*, \eta_i^*)\) under an optimal mechanism have the following properties:

1. \(d_i(s_i^*)\eta_i^*\) is nondecreasing in \(i\).
2. Payer’s benefit is decreasing in \(b_i\), but IR constraints for \(i \neq 1\) are not necessarily tight.
3. All IR and IC constraints are satisfied and payer’s benefit is maximized when one-step IC constraints are tight and the lowest type makes zero profit.
4. \(s_i^*\) is increasing in \(i\).
5. If \(\nu_s = 0\), i.e., proposers’ demands are independent, then every proposer type is selected with probability 1.
6. If \(\gamma > \alpha\), \(\mu \geq 2\bar{s}_i(\ell)\nu_s(n - 1)\) and \(\lambda \geq \nu_s(n - 1) - \bar{s}_i(\ell)/\bar{s}_i(1)\), then every proposer type is selected with probability 1, and \(s_i^* = \bar{s}_i(i)\) and \(b_i^* = b^h(i)\).

From Proposition 8, we can also show that

\[ b_i^* = \text{zero-profit payment}_{i-s} \]

+ informational rent or external subsidy

\[ = (\alpha s_i^* + \beta_i + \epsilon_i) \]

\[ + \sum_{k=1}^{\ell - 1} (\beta_i + \epsilon_i) - (\beta_{i+k} + \epsilon_{i+k}) d(s_i^*)\eta_i^* \]

\[ = \frac{d(s_i^*)\eta_i^*}{d(s_i^*)\eta_i^*} (C1) \]

For a sequence of quality scores, \(\sum_{i=1}^{\ell} (\beta_i + \epsilon_i) - (\beta_{i+k} + \epsilon_{i+k}) d(s_i^*)\eta_i^* / d(s_i^*)\eta_i^*\) is the size of the external subsidy needed for incentivistic compatibility. By setting \(s_i^* = s_i^*\), we can also obtain the size of the subsidy needed to realize a first-best solution (see Myerson and Satterthwaite 1983 for a general result proving the corresponding impossibility theorem). We also obtain the following insights by comparing the first-best solution and the solution under the optimal mechanism.

Proposition 9. The following statements are true when both first-best solution and optimal mechanism select type-i proposer (i.e., \(0 < \eta_i^* \leq 1\) and \(0 < \eta_i^* \leq 1\)):

1. If \(b_i^* = b_i^h\) and \(b_i^* = b_i^h\), then \(s_i^* \leq s_i^* = \bar{s}_i(i)\).
2. If \(b_i^* < b_i^h\) and \(b_i^* > b_i^h\), then \(s_i^* > s_i^*\).
3. If \(b_i^* < b_i^h\) and \(b_i^* = b_i^h\), then both \(s_i^*\) and \(s_i^*\) are less than \(\bar{s}_i(i)\). Moreover, we can have either \(s_i^* < s_i^*\) or \(s_i^* > s_i^*\).

Endnotes

1. The term payer refers to an entity that is contractually obligated to pay for medical services consumed by a group of people that we refer to as beneficiaries.
2. The following weights have been announced: service model design (20); financial model (40); quality of care and patient centeredness (25); and organizational capabilities, prior experience, and readiness (15); see Center for Medicare and Medicaid Innovation (2011).
3. There is some evidence in the literature that competition could lead to worsening quality; see e.g., Decarolis (2014).
4. In practice, CMS tracks several self-reported quality metrics, some of which relate to all beneficiaries served by proposers, and some others only to those that enroll in the bundled services. Furthermore, proposer’s organizational capabilities, service design, and financial arrangements among participating providers also play a role in proposer selection. Because each components’ weights are specified by CMS, there is no loss of generality in combining all metrics not related to target payments into a single scalar metric that we call quality.

Proposers that have been selected so far for the prospective payment model seem to have a dominant presence in their markets for the proposed DRG either because of their existing reputation or because they managed to put together a large network of providers. This suggests that organizations proposing bundles tend to have significant organizational capability. We also find that the specification of quality metrics is in a state of flux. CMS is still fine tuning which metrics need to be reported. Also, several metrics pertain to the absence of undesirable events with a focus on avoiding poor outcomes, rather than improving positive outcomes.

5. Leakage depends mostly on the size of service providers’ network included in a proposer’s service chain. Therefore, we model leakage as a type-dependent private attribute of each proposer.
6. Note that our setting is also different from the affiliated private values and common values environments. In the former, the signals themselves may be correlated, but each proposer, given its signal, believes its private costs to be accurate. In contrast, in common values environment, proposers’ costs are correlated and no single proposer’s individual signal is sufficient to accurately estimate the true cost. Both these environments give rise to a phenomenon known as the winner’s curse, which is absent in our setting.
7. We justify the assumption of symmetry using two lines of reasoning. First, proposers do not know the cost functions of other proposers because proposing entities are formed in response to the request for applications by CMS. Therefore, a proposer has no particular information advantage over others, which justifies
the assumption of a common distribution of types. The second line of reasoning is rooted in analytical convenience. In particular, without symmetry, no structural results are possible because depending on what one assumes about the nature of heterogeneity, one can obtain a whole host of different outcomes. Our goal in this paper is to show that uncertain mechanism could be better than fixed threshold. We do so in the environment in which proposers are symmetric, which serves our purpose in this paper.

8. As stated in Edmunds and Sloan (2012, p. 1) “The Medicare system adjusts fee-for-service payments to hospitals and practitioners according to the geographic location in which providers practice, recognizing that certain costs beyond providers’ control vary between metropolitan and nonmetropolitan areas and also differ by region. The fundamental rationale for geographic adjustment is to create a payment structure that adjusts payments for input price differences that healthcare professionals and institutions face, such as the cost of employee compensation.” When combined with the fact that most people seek medical services from local providers, the implication for our setting is that localities used by CMS to adjust prices may be treated as separate markets.

9. The benefit per proposer is the product of its equilibrium score and the number of beneficiaries it attracts. Also, in this environment, every Medicare beneficiary is covered by CMS regardless of whether he or she obtains services. This includes services offered by all qualified medical service providers, whether individual practitioners or institutions, irrespective of their status as bundled providers or not. Furthermore, service providers determine which services are needed, e.g., the number of presurgery office visits required by a surgeon, and CMS currently pays on a fee-for-service basis. That is, the payment made on behalf of a beneficiary could be quite different even in the same geographical area depending on the beneficiary’s choice of service providers.

10. We highlight a specific feature of procurement mechanisms in healthcare setting with the help of an example. Consider two proposers with historical average payments for an identical bundle of $100 and $1,000, respectively. Suppose both offer the same quality score and the number of beneficiaries it attracts. Also, in this environment, every Medicare beneficiary is covered by CMS regardless of whether he or she obtains services. This includes services offered by all qualified medical service providers, whether individual practitioners or institutions, irrespective of their status as bundled providers or not. Furthermore, service providers determine which services are needed, e.g., the number of presurgery office visits required by a surgeon, and CMS currently pays on a fee-for-service basis. That is, the payment made on behalf of a beneficiary could be quite different even in the same geographical area depending on the beneficiary’s choice of service providers.

References


Diwakar Gupta is a professor of industrial and systems engineering at the University of Minnesota. He also serves as the Program Director for Service, Manufacturing, and Operations Research program at the National Science Foundation. Diwakar’s research focuses on healthcare delivery systems, state transportation agencies’ operations, and supply chain and revenue management.

Mili Mehrotra is an assistant professor of supply chain and operations in the Carlson School of Management at the University of Minnesota. Her research lies in the domain of socially responsible supply chains and operations management. In particular, she is interested in developing and analyzing incentive schemes, and studying coordination and optimization issues that arise in practice because of the actions taken by a wide variety of stakeholders for achieving broader social objectives.