Optimizing Healthcare Network Design under Reference Pricing and Parameter Uncertainty

Victoire Denoyel[a], Laurent Alfandari[b,*], Aurélie Thiele[c]

[a] *ESSEC Business School, avenue Bernard Hirsch, Cergy-Pontoise, France*
[b] *ESSEC Business School, avenue Bernard Hirsch, Cergy-Pontoise, France*
[c] *SMU, 3145 Dyer St, Dallas, TX, USA*

Abstract

Healthcare payers are exploring cost-containing policies to steer patients, through qualified information and financial incentives, towards providers offering the best value. With Reference Pricing (RP), a payer determines a maximum amount paid for a procedure, and patients selecting a provider charging more pay the difference. In a Tiered Network (TN), providers are stratified according to criteria such as quality and cost, each tier having a different out-of-pocket price. Motivated by a program recently implemented in California, we design two optimization models for payers combining both RP and TN, filling the gap of quantitative research on these novel payment policies. The main decision is to select which providers to exempt from RP, whose patients will face no out-of-pocket costs. The objective is to minimize the payer's cost, while constraints provide decision makers with levers for a trade-off between cost reduction and patients' satisfaction. We build robust counterparts for both models to account for parameter uncertainty. Numerical experiments provide insights into how tiers are scattered on a price/quality plane. We argue that this system has strong potential in terms of costs reduction, quality increase for patients and visibility for high-value providers.

Keywords: Combinatorial optimization, OR in health services, Robust optimization, Choice models, Fractional optimization'

1. Introduction

In the United States, healthcare providers can charge very different prices for the same procedure. A study by Hsia et al. (2012) reports prices ranging from \$1,529 to \$182,955 for an appendectomy, although no clear difference in care quality could account for this range. The Health Care Cost Institute states in a 2014 report: "Rising prices, rather than utilization, were the primary drivers of spending growth for all medical service categories and brand prescriptions" (Health Care Cost Institute, 2014). In the fee-for-service system, patients are not incentivized to choose the best value provider, and providers are not incentivized to be cost-efficient; but as an indirect consequence, costs for payers and premiums for patients are increasing.

*Corresponding author
Email address: alfandari@essec.edu (Laurent Alfandari)

One of the emerging tools for payers to protect themselves and patients from this high-price spiral is Reference Pricing (RP). In RP, the payer's liability is capped to a predefined amount, hereby referred to as "reference price". Patients are asked to pay the difference between the posted price and the reference price, if there is one. This system has the potential to increase value for all stakeholders: for payers, it can help containing the price rise. It can steer patients towards providers with the highest quality/price ratio. It can shift patients' attention towards their care quality, and help them make more price-conscious choices. It can also improve cooperation between payers and providers, and allow more visibility for the best performing providers. Yet there are important caveats: RP needs price transparency as well as reliable data regarding providers quality. It is also not applicable to any kind of health procedure: emergency or routine procedures are out of the scope. Implementation programs should be focused on specific procedures with a large and volatile price range. It also has the inconvenient that providers could redistribute the effect of the reimbursement decrease on other procedures with less control.

Another cost-containing measure is for payers to design Tiered Networks (TN), also known as tiered plans or tiered products. Providers are grouped in tiers based on criteria such as price, quality or location. Patients are subject to different co-payments according to the tier of provider they visit. Emanuel et al. (2012) strongly advocate the implementation of tiered plans for payers.

Inspired by a pilot experiment described in the next section, we provide the first decision model for the combination of RP and TN for a healthcare procedure, including the choice of providers subjected to this payment scheme instead of fee-for-service. Our main contribution is to present a methodological framework that captures the impact of quality, volume and out-of-pocket payment on patient usage of specific facilities. The specificities of this framework, namely a choice model where demand depends on binary decision variables, as well as the linearization of the resulting fractional model, are also a novelty for healthcare optimization problems.

The following section contains a literature review for different research streams to which this paper contributes. Section 3 includes a presentation of two models with analytical insights. Section 4 consists of numerical experiments and Section 5 discusses how to address parameter uncertainty. Section 6 concludes the paper.

2. Literature Review

2.1. Reference Pricing: brief history and relevant literature

RP is not new to the healthcare industry and quite used in the pharmaceutical sector: it was introduced in Germany in 1989 with the Statutory Health Insurance System, and later in many European and Commonwealth countries (Brekke et al., 2007). An abundant stream of empirical literature exists on the impact of RP on drugs prices and competition. In a broad cross-country study, Danzon and Ketcham (2003) argue that the goals of RP (such as encouraging price competition) are not achieved in pharmaceutical markets; López-Casasnovas and Puig-Junoy (2000) however report "indications of a greater price competition" as the result of RP. Important parameters such as the supply structure and the price ranges differentiate the pharmaceutical industry from the hospital industry.

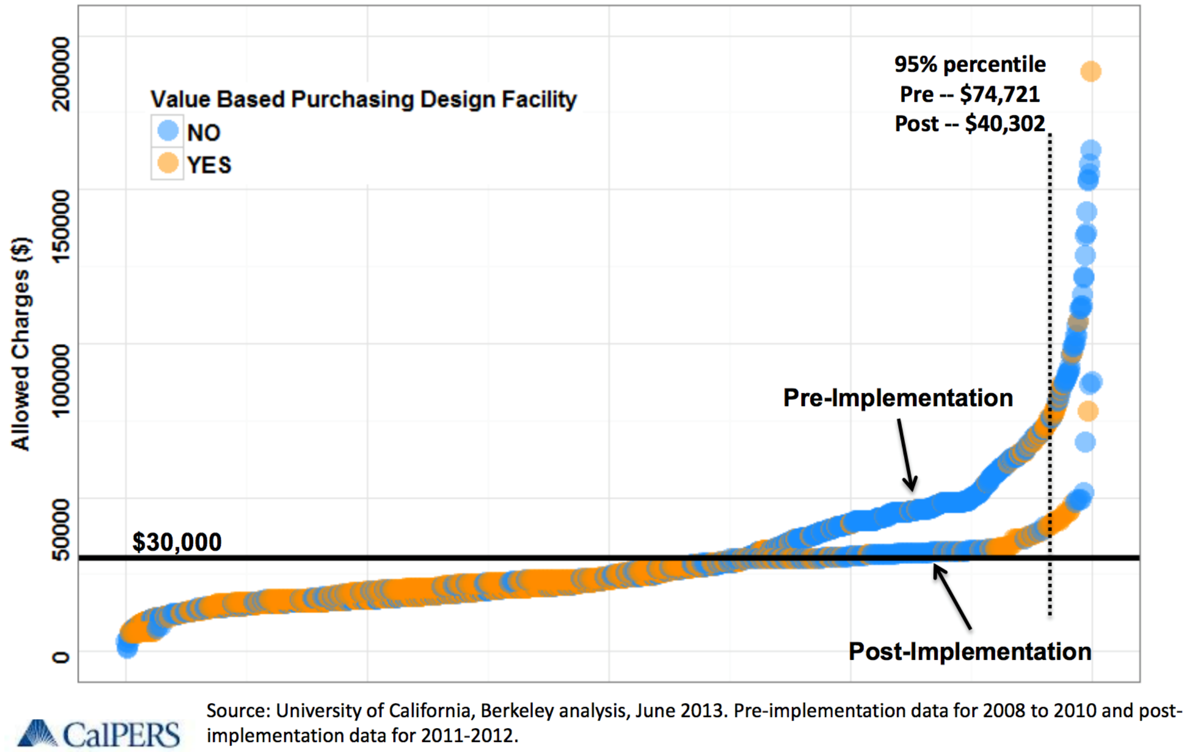
In the context of healthcare procedures, RP was modeled by Enthoven in 1977 under the name "Consumer Choice Health Plan", as quoted in Reinhardt (2013), and yet widely neglected since then. Health economist Uwe Reinhardt recently called RP the "Sleeper" of the healthcare reform, suggesting by this image both an untapped resource and a potential key element (Reinhardt, 2013).

Recent implementation by Anthem Blue Cross for the California Public Employees' Retirement System (CalPERS) generated a healthy debate about RP merits and legitimacy. Notably, the scheme designed by the organization exempts some of the providers in the network (designated as "Value-Based Purchasing Design" facilities) from the RP obligation. One explicit goal is to favor high-value providers. The experiment was applied specifically to hip or knee replacement procedures. Robinson and Brown (2013) evaluated the impact of this implementation on the demand allocation between facilities as well as on prices. They found that patients demand shifted from providers subject to RP towards providers exempted from RP, especially from high-price providers to low-price providers, and that overall prices were reduced by 20.2% on average (see Figure 1).

This suggests that RP is capable of influencing providers' prices and patients' choices. CalPERS extended this program to cataract removal surgery, as studied in Robinson et al. (2015): more specifically, this program was aimed at redirecting patients towards ambulatory centers rather than hospital departments, through a reference pricing design. The authors outlined a 21.1% demand shift towards ambulatory centers and a 38.5% price reduction in two years for this procedure. These findings confirm the potential in increasing patients financial responsibility. However, for the payer as a decision maker, specific questions remain unanswered:

- How to decide which providers are to be exempted from RP?
- How to incorporate quality or value instead of cost only in her decisions?
- How to set an adequate reference price?

Figure 1: CalPERS study: price repartition and selection of providers (Cowling, 2013)



To the best of our knowledge, there currently exists no formal optimization model aiding the payer to make those decisions. We fill this gap by creating a model oriented towards practitioners, inspired to a large extent from the CalPERS implementation.

2.2. Theoretical Contributions

While operations research has been extensively used for healthcare aspects such as planning or logistics, it is less used in the pricing and policy aspects (see the surveys by Brailsford and Vissers, 2011 or Rais and Viana, 2011). In a recent contribution, Andritsos and Tang (2014) use a queuing framework to analyse the implication from competition in European healthcare services. From a modeling standpoint, our problem has important similarities with project portfolio selection problems, if providers are seen as "projects" to select while minimizing a cost function and respecting some criteria thresholds. Project portfolio selection problems have been extensively studied in the literature; Heidenberger and Stummer (1999) reviews quantitative modeling approaches in that field. Similarly, another relevant branch of the traditional operations management literature is supplier selection (see Ho et al., 2010 for a recent literature review). The following aspects make both our quantitative formulation and our solution approach novel:

- Because we assume that providers and patients will adjust their decisions to the network structure (the providers who are exempted and those who are subject to RP), both cost and volume for each "portfolio" item (provider) depend on the global composition of the network. To the best of our knowledge, our model is the first to include this particular setting.

- The choice model we develop, based on the classical Multinomial Logit (MNL) model (McFadden, 2001), leads to a linear-fractional formulation in our context of provider selection/exemption from the RP scheme, which we show how to solve efficiently for the problem at hand.
- The need to incorporate constraints on patient satisfaction and network quality makes the analysis of this problem fundamentally different from traditional R&D project management or supply chain management, and specific to the healthcare policy sector.
- We derive a threshold policy (in Theorem 3.8) describing how to select the network when there are no quality and dissatisfaction constraints. Starting from that network, the way these constraints are configured can be considered as a lever allowing the decision-maker to strategically shape the network selection.
- In order to decrease the risk due to parameter uncertainty, we develop robust counterparts for both models, using the Bertsimas and Sim approach (Bertsimas and Sim (2004)). Using robust optimization for choice models is an important theoretical contribution in itself. On a similar line, Rusmevichientong and Topaloglu (2012) developed static and dynamic robust optimization models under the MNL choice model for revenue management applications.

Let us note that several approaches exist for Robust Optimization (see the recent survey of Gabrel et al., 2014). For example, Taguchi-based approaches use simulations and aim at optimizing an output by regression of the variables on parameters than can vary according to a given probabilistic distribution. In Dellino et al. (2010), some parameters follow a normal distribution and the expected value of the output is optimized while constraining the variance to be below a given threshold (see also Dellino et al., 2012). When parameter distributions are unknown and problem constraints prevent from using regression techniques, e.g. for Linear Programming problems, the deterministic approach of Bertsimas and Sim (2004) has become a classical way to deal with uncertainty. In this approach, uncertainty is modeled by a variation interval without assuming to know the distribution of the uncertain parameter. This approach does not use simulation, but optimizes the worst case over all possible parameter variations in a deterministic way. It extends the work of Ben-Tal and Nemirovski (1999) and the (more conservative) original approach of Soyster (1973) to controlled uncertainty, where only a limited number of parameters is allowed to vary. The success of this approach, that uses duality, partly lies in the fact that the robust counterpart of a Linear Program remains a Linear Program, and therefore it is generally computationally tractable. Finally, observe that the Bertsimas and Sim (2004) paper showed empirically on benchmark instances that even if it focuses on worst-case optimization, it is empirically not too conservative, i.e., the average loss of value, when adding protection against parameter variation to ensure feasibility, is not large.

In the following section, we present the two deterministic models.

3. Optimization Models for Healthcare Network Design

We consider an ecosystem composed of one payer (the decision-maker) and her network of providers. We isolate a single procedure, characterized by a large price value and volatility. Each provider is characterized, regarding this specific procedure, with his own posted price, volume of patients (or market share), and quality measure. The payer observes this set of parameters and decides which providers to exempt from RP, and which ones to set on an RP contract. The reference price is determined separately based on negotiations and design requirements such as the percentage of patients in the network who will not incur any out-of-pocket payments.

The payer minimizes her expected cost for this particular procedure. Our models specifically take into account the consequences of her decisions, namely, the general decrease in prices and the redistribution of patient demand between providers. We consider two models:

1. A *homogeneous model* (Model 1), where all providers who are exempted from RP see a relative increase of $\beta_1 \in (0,1)$ (given, independent of a provider's characteristics) in patient volume. Providers who are subject to RP then see a relative decrease in patient volume that is also independent of the provider, and is computed so that total patient volume is preserved. This model isolates the effect of the RP implementation on patients choice of a provider, excluding other choice factors. It is particularly well-suited when little information about patient choice is available.
2. A *heterogeneous model* (Model 2), where the change in patient volume for each provider following the RP implementation depends on that provider's own characteristics (such as quality or out-of-pocket payments faced by the patient) through a Multinomial Logit Model, which also conserves total volume. This model is slightly more complicated but is a more realistic representation of reality, when the appropriate input data is available.

Although it may look simplistic to assume an homogeneous increase rate for exempted providers in Model 1, the interest of this model also lies in its robust counterpart, described in section 5. Assume that in average the percentage of volume increase of exempted providers is 20%, like in the CalPERS study. If we know the typical variation over all providers is in $[10\%, 30\%]$, we can plug this uncertainty interval into the robust model. The worst-case variation will then cover all possible heterogeneous variations of providers inside the interval.

For each model, we seek to minimize total payer cost subject to constraints on (a) quality (the average quality of exempted providers must be superior to the average provider quality for the whole network to justify the exemption from RP) and (b) patient satisfaction (patients are dissatisfied when their pre-implementation provider is not exempted from RP, especially if he is of high quality, and the total number of dissatisfied patients cannot exceed a given threshold).

Table 1 summarizes the notations we use in the remainder of the paper.

3.1. Model 1: Homogeneous Model

3.1.1. Assumptions on volume, price and quality:

Volumes. We consider a short-term decision framework (one or two years), which is not enough to look at patients' market decisions. As a consequence, all patients stay in the system, that is, the total volume is preserved. The volume for each provider can vary in two ways:

- If provider i is exempted from RP ($y_i = 1$), his initial volume v_i is anticipated to increase, both because of the out-of-pocket price decrease and of the "stamp of approval" from the payer that it represents. Let β_1 be the relative volume increase for exempted providers: the volume at provider i after implementation is $(1 + \beta_1)v_i$.

- If provider i is subject to RP ($y_i = 0$), his volume is anticipated to decrease for the same reasons, due to total volume conservation. Let β_2 be the relative volume decrease for non-exempted providers. The anticipated volume at provider i after implementation is $(1 - \beta_2)v_i$.

Making the assumption that the network design problem is not trivial (there exists at least one i for whom $y_i = 0$), volume conservation entails:

$$\beta_2 = \beta_1 \frac{\sum_{i \in I} v_i y_i}{\sum_{i \in I} v_i (1 - y_i)}.$$

Table 1: Notations: Sets and Parameters

I	set of n providers ($n \geq 1$)
	Volumes
v_i	observed volume at provider i before RP implementation
V	total initial volume
$\tilde{v}_i$	anticipated volume at provider i after RP implementation
β_1	relative increase in volume in Model 1 for providers exempted from RP ($\beta_1 \in [0,1]$)
	Prices
p_i	observed price at provider i before RP implementation
$\bar{p}$	reference price
p_i^+	surplus of initial price over reference price for provider i
$\mathring{p}_i$	anticipated price charged by provider i
$\tilde{p}_i$	anticipated price paid by payer for provider i
p'_i	anticipated out-of-pocket amount for patient at provider i
f	function of the surplus of initial price over reference price after RP implementation
	Choice model parameters
a	Weight of out-of-pocket price
b	Weight of quality indicator
c	Constant
d	Weight of the inclusion in the network
	Constraints
q_i	quality measure for provider i
$\bar{q}$	average quality over all providers
π_i	or $\pi_i(q_i, p_i)$: probability that patients of provider i are dissatisfied if i is subject to RP
μ	upper bound for the proportion of dissatisfied patients
α	average quality improvement for the set of providers exempted from RP ($\alpha \in [0,1]$)
	Decision Variables
y_i	$= 1$ if provider i is exempted from RP, 0 otherwise
β_2	volume decrease rate for providers subject to RP (Model 1)
w_i	$= \beta_2$ if $y_i = 0$ and 0 otherwise (Model 1).

Prices. We expect a significant decrease in prices that are above the reference price, as Figure 1 illustrates. Table 2 summarizes possible scenarios and the cost sharing between payer and patient.

Table 2: Price impact

		$p_i \leq \bar{p}$	$p_i \geq \bar{p}$
New price $\mathring{p}_i$		p_i	$\bar{p} + f(p_i - \bar{p})$
Payer part $\tilde{p}_i$	$y_i = 0$	p_i	$\bar{p}$
	$y_i = 1$	p_i	$\bar{p} + f(p_i - \bar{p})$
Patient part p'_i	$y_i = 0$	0	$f(p_i - \bar{p})$
	$y_i = 1$	0	0

If a provider's price before RP implementation falls below the reference price, we assume that he will neither decrease or increase it after RP implementation because patients continue to have no out-of-pocket expenses beyond deductibles. Deductibles would not depend on the chosen provider and thus are omitted. If a provider's price before RP implementation exceeds the reference price, we assume that it will decrease after RP implementation.

Specifically, the latter will partly reduce their surplus above the RP. Let f be the function quantifying the excess price that remains after the prices are adjusted post RP-implementation. For analytical insights we make the following assumptions: f is continuous, increasing and concave in $(p_i - \bar{p})$: the larger the gap between a price and the reference price, the higher the decrease. f is equal to 0 over $(-\infty, 0]$, that is, when $p_i \leq \bar{p}$, which allows continuity between the two columns of Table 2. Besides, we assume that $f(x)$ belongs to $[0, x]$ for all x: a provider cannot decrease her price below the reference price nor increase it above her initial price. Finally, we assume that $df(p_i^+)/d\bar{p} \geq -1$ (on any interval where $f(p_i^+)$ is differentiable), which means that the marginal decrease in the price post-implementation cannot be higher than the marginal decrease in the gap between a price pre-implementation and the reference price. To implement the model without the analytical insights, f simply needs to be continuous increasing and $f(x)$ belongs to $[0, x]$ for all x.

Next, $\tilde{p}_i$ represents the part of the price paid by the insurer (payer). Below the reference price, the payer pays the full price. Above the reference price, if a provider is subject to RP ($y_i = 0$), the amount paid by the insurer is capped to $\bar{p}$; if a provider is exempted from RP ($y_i = 1$), the full price is paid by the insurer (payer). Finally, p'_i represents the out-of-pocket patient price, which is the difference between the new price and what the insurer pays. It is only positive for a provider who is subject to RP with a price higher than the reference price. By construction, $\mathring{p} = \tilde{p} + p'$.

The charge repartition between the payer and the patient can therefore be summarized as:

$$\tilde{p}_i = \min(p_i, \bar{p}) + y_i f(p_i^+) \qquad \forall i \in I$$

$$p'_i = (1 - y_i) f(p_i^+) \qquad \forall i \in I$$

where $p_i^+ = max(p_i - \bar{p}, 0)$.

Remark. Without loss of generality, this model can include copayment or coinsurance terms. If K is the fixed amount due as copayment, we simply replace prices p_i by $p_i - K$. If τ is the percentage due as coinsurance, the price should be replaced instead by $(1-\tau)p_i$.

Quality. We capture the quality of a provider in a one-dimensional measure q_i. A higher q_i means a higher quality of the provider.

3.1.2. Formulation:

We minimize the total payer cost, i.e., the sum over providers of their anticipated volume multiplied by the anticipated payment they receive from the payer. The expression of the anticipated volume depends on whether the provider is exempted from RP ($y_i = 1$) or subject to RP ($y_i = 0$).

$$obj = \sum_{i \in I} \underbrace{[(1+\beta_1)v_i y_i + (1-\beta_2)v_i(1-y_i)]}_{\text{anticipated volume}} \underbrace{(\min(p_i, \bar{p}) + f(p_i^+)y_i)}_{\text{anticipated charges}}$$

We introduce three constraints in Model $(M1a)$ below.

- Eq. (2) ensures that the subset of exempted providers have a "better-than-average" quality than the network as a whole. This was validated by a discussion with an expert who studied RP-exemption networks, and was consistent with the CalPERS study. The average quality of exempted (or in-network) providers should improve by a percentage α the average quality of the initial set, where parameter α is a choice of the decision-maker. Another modeling possibility could be that each selected provider should have a quality above a given threshold (say, a rating of 3 out of 5). However, this would exclude many providers with low prices that enable to cover a larger market, which is why the quality requirement is on average. The expert confirmed that the RP-exemption networks also included low-quality providers. Indeed, we consider a general insurer who provides a portfolio of providers diversified enough to attract all segments of the population. The way the constraint is modeled allows us to insure that if lower-quality providers are included in the network, their price is also on the lower end.

- Eq. (3) controls the risk of patients leaving the insurer network in the mid-term. It states that the percentage of dissatisfied patients should not exceed a given control parameter μ, which is also a choice of the decision-maker. A given proportion π_i of pre-implementation patients of provider i will be dissatisfied if i is subject to RP; this proportion is assumed to increase with the provider's quality and out-of-pocket price.

- Eq. (4) ensures that the total volume of patients is conserved. This is a short-term property, provided that dissatisfied customers are likely to transform into lost customers for the insurer in the mid-term.

This yields the following nonlinear model:

$$(M1a)\ \min_{y,\beta_2} \sum_{i\in I} v_i[\min(p_i,\bar{p})(1-\beta_2)+(\min(p_i,\bar{p})(\beta_1+\beta_2)+f(p_i^+)(1+\beta_1))y_i] \tag{1}$$

$$s.t \quad \sum_{i\in I} q_i y_i \geq (1+\alpha)\bar{q}\sum_{i\in I} y_i \tag{2}$$

$$\sum_{i\in I} \pi_i v_i(1-y_i) \leq \mu V \tag{3}$$

$$\beta_2 \sum_{i\in I} v_i(1-y_i) = \beta_1 \sum_{i\in I} v_i y_i \tag{4}$$

$$\beta_2 \in [0,1], \quad y_i \in \{0,1\} \qquad \forall i \in I \tag{5}$$

Both constraint (4) and the objective function are nonconvex quadratic. Theorem 3.1 shows that Model ($M1a$) can be reformulated as a MIP and thus be solved more efficiently.

Theorem 3.1 (MIP formulation of Homogeneous Model)**.** *Model* ($M1a$) *is equivalent to the MIP linear model:*

$$(M1)\ \min_{y,w,\beta_2} \quad z_1 = \sum_{i\in I}\left[v_i \min(p_i,\bar{p})(1-w_i) + \left(\min(p_i,\bar{p})\beta_1 + f(p_i^+)(1+\beta_1)\right) v_i y_i\right] \tag{6}$$

$$s.t \quad \sum_{i\in I}(q_i-(1+\alpha)\bar{q})y_i \geq 0 \tag{7}$$

$$\sum_{i\in I} \pi_i v_i y_i \geq \sum_{i\in I} \pi_i v_i - \mu V \tag{8}$$

$$\sum_{i\in I} v_i(\beta_1 y_i - w_i) = 0 \tag{9}$$

$$\beta_2 - y_i \leq w_i \leq \beta_2 \qquad \forall i \in I \tag{10}$$

$$w_i \leq 1 - y_i \qquad \forall i \in I \tag{11}$$

$$y_i \in \{0,1\}; w_i, \beta_2 \in [0,1] \qquad \forall i \in I \tag{12}$$

Proof. Follows directly from introducing a set of auxiliary variables w_i and the two sets of constraints (10) and (11) to ensure that, for all i, w_i is equal to β_2 if $y_i = 0$ and 0 otherwise. □

3.1.3. Insights into the optimal strategy.

We refer to the ideal strategy as the optimal strategy without the quality and satisfaction constraints: ($M1a$) with Eqs. (1), (4)-(5), or ($M1$) with Eqs. (6), (9)-(12).

Theorem 3.2. *(i) The ideal strategy is to exempt from RP all providers whose price is at most $\bar{p}$ and not exempt from RP all providers whose price exceeds $\bar{p}$.*
(ii) It is optimal for the decision-maker solving ($M1$), i.e., the problem with quality and satisfaction constraints, to exempt from RP all providers whose price is at most $\bar{p}$ and quality is at least $(1+\alpha)\bar{q}$.

Proof. We study the change in the objective if y_l goes from 0 to 1 for some l and the other binary decision variables remain unchanged. We denote β_2^o and β_2^n, respectively, the old value of β_2 when $y_l = 0$ and the new value after the change, when $y_l = 1$. From Eq. (1), we have:

$$\Delta obj = v_l \left[\min(p_l, \bar{p})(\beta_1 + \beta_2^n) + f(p_l^+)(1+\beta_1)\right] + \left(v_l \min(p_l, \bar{p}) + \sum_{i \neq l} v_i \min(p_i, \bar{p})(1 - y_i) \right) (\beta_2^o - \beta_2^n). \tag{13}$$

We can also write:

$$\beta_2^o = \beta_1 \frac{\sum_{i \neq l} v_i y_i}{V - \sum_{i \neq l} v_i y_i}, \text{ and } \beta_2^n = \beta_1 \frac{v_l + \sum_{i \neq l} v_i y_i}{V - \sum_{i \neq l} v_i y_i - v_l},$$

so that:

$$\beta_2^o - \beta_2^n = -V\beta_1 \frac{v_l}{\left(V - \sum_{i \neq l} v_i y_i\right)\left(V - \sum_{i \neq l} v_i y_i - v_l\right)}. \tag{14}$$

Note also that:

$$\beta_1 + \beta_2^n = \beta_1 \frac{V}{V - \sum_{i \neq l} v_i y_i - v_l}. \tag{15}$$

If $p_l \leq \bar{p}$, we reinject $\min(p_l, \bar{p}) = p_l$ and $f(p_l^+) = 0$ into Eq. (13) as well as Eqs. (14)-(15) to obtain:

$$\Delta obj = -\frac{v_l \beta_1 V}{V - \sum_{i \neq l} v_i y_i - v_l} \cdot \frac{\sum_{i \neq l} v_i \{\min(p_i, \bar{p}) - p_l\}(1 - y_i)}{v_l + \sum_{i \neq l} v_i (1 - y_i)}.$$

Start from the zero vector, and assume w.l.o.g. that the providers have been numbered in increasing order of prices, breaking ties arbitrarily. Then switching y_1 from 0 to 1 decreases the objective, since $\min(p_i, \bar{p}) \geq p_1$. Now, assume that there exists j such that only providers 1 to j are exempted, with $p_{j+1} \leq \bar{p}$. Then p_{j+1} is the smallest of all prices among non-exempted providers, so that $\sum_{i=j+2}^n v_i(\min(p_i, \bar{p}) - p_{j+1})(1 - y_i) \geq 0$ and switching y_{j+1} to 1 decreases the objective. This allows us to conclude that it is optimal in the ideal strategy to exempt providers whose price is at most $\bar{p}$, and in the provider-selection problem $(M1)$ to exempt providers whose price is at most $\bar{p}$ and whose quality is at least $(1+\alpha)\bar{q}$, since both the objective and the constraints are no worse. Note that this is a sufficient but not necessary condition for providers to be exempted in $(M1)$. Some providers with quality less than $(1+\alpha)\bar{q}$ may also be exempted at optimality in $(M1)$. The ideal solution does not have the quality requirements.

To prove that in the ideal strategy, it is optimal not to exempt providers whose price exceeds $\bar{p}$, start from the strategy obtained with the previous procedure, with all providers charging at most $\bar{p}$ exempted and the others not exempted. Let l be such that $p_l > \bar{p}$. Eq. (13) becomes:

$$\Delta obj = v_l \left[\bar{p}(\beta_1 + \beta_2^n) + f(p_l^+)(1+\beta_1)\right] + \left(v_l \bar{p} + \sum_{i \neq l} v_i \min(p_i, \bar{p})(1 - y_i) \right) (\beta_2^o - \beta_2^n).$$

But because all non-exempted providers have a price higher than the RP, $\sum_{i \neq l} v_i \min(p_i, \bar{p})(1 - y_i) = \bar{p} \sum_{i \neq l} v_i (1 - y_i)$. This yields:

$$\Delta obj = v_l f(p_l^+)(1+\beta_1) + \bar{p} \left[v_l(\beta_1 + \beta_2^n) + (v_l + \sum_{i \neq l} v_i (1 - y_i)) (\beta_2^o - \beta_2^n) \right].$$

Using $v_l + \sum_{i \neq l} v_i(1 - y_i) = V - \sum_{i \neq l} v_i y_i$ and Eqs. (14)-(15), we see that the term between the straight brackets is equal to 0, so that $\Delta obj = v_l f(p_l^+)(1 + \beta_1)$, which is always positive. Hence, it is never optimal in the ideal strategy to exempt a provider charging more than the RP. □

In this model, the shift in volumes is independent of an important factor such as the out-of-pocket price paid by the patient, which may be quite restrictive in a real-life context. This leads us to develop Model 2, where the increase/decrease rates of the volumes at the provider level depend on both data (prices) and decisions (in-network selection).

3.2. Model 2: Heterogeneous Model

In this section, we develop an alternative model where we build upon a more fine-grained relationship between out-of-pocket payment and procedure volume at the provider level. We model the cost as: $z_2 = \sum_{i \in I} \tilde{v}_i \tilde{p}_i$ where $\tilde{v}_i$ represents the anticipated volume after the implementation of RP, and $\tilde{p}_i$ the anticipated price charged to the payer. The latter is modeled exactly as in Model 1 (see Table 2). We explain the modeling of volumes in detail below.

3.2.1. Modeling:

Volume impact. We represent the anticipated volume of providers with a MNL choice model (McFadden, 2001). The probability for patients to choose a provider i depends on their utility for it, a linear function of the out-of-pocket price p'_i, as well as on the quality of the provider and his selection by the payer. The anticipated volume for provider i is:

$$\tilde{v}_i = V \frac{e^{-ap'_i + bq_i + c + dy_i}}{\sum_{k \in I} e^{-ap'_k + bq_k + c + dy_k}} \qquad \forall i \in I$$

where V is the total patients volume, and (a, b, c, d) are positive choice parameters.

The total volume of patients is conserved, and the volume of a provider is lower for a higher out-of-pocket price, higher for a better quality, and higher for a selected provider. Note that in the initial case (before RP implementation), we obtain by setting all out-of-pocket prices and y_i to 0:

$$v_i = V \frac{e^{bq_i + c}}{\sum_{k \in I} e^{bq_k + c}} \qquad \forall i \in I$$

The validity of this expression relies on two assumptions: first, as mentioned earlier, the initial co-payment is equal for all providers (since it is set by the payer), which allows us to disregard it in the provider selection problem. The second assumption is that the implementation of the RP scheme does not affect problem parameters such as the quality q_i and the choice model parameters. Consequently, the expression of volume can be reformulated as:

$$\tilde{v}_i = V \frac{v_i e^{-ap'_i + dy_i}}{\sum_{k \in I} v_k e^{-ap'_k + dy_k}} = V \frac{v_i e^{-af(p_i^+)(1-y_i) + dy_i}}{\sum_{k \in I} v_k e^{-af(p_k^+)(1-y_k) + dy_k}} \qquad \forall i \in I \tag{16}$$

A key insight of Eq. (16) is that the decision-maker does not have to estimate parameters b and c; the information she needs is captured in the v_i. Only the a and d need to be estimated.

Cost function. After injecting the expressions for p_i' and $\tilde{p}_i$, we write the cost function as:

$$\begin{aligned}
z_2 &= \sum_{i\in I} \tilde{v}_i \tilde{p}_i \\
&= V \sum_{i\in I} \frac{v_i e^{-af(p_i^+)(1-y_i)+dy_i}}{\sum_{k\in I} v_k e^{-af(p_k^+)(1-y_k)+dy_k}} (\min(p_i,\bar{p}) + f(p_i^+)y_i) \\
&= V \sum_{i\in I} \frac{v_i e^d y_i + v_i e^{-af(p_i^+)}(1-y_i)}{\sum_{k\in I}(v_k e^d y_k + v_k e^{-af(p_k^+)}(1-y_k))} (\min(p_i,\bar{p}) + f(p_i^+)y_i) \\
&= V \frac{\sum_{i\in I} v_i [e^d(\min(p_i,\bar{p}) + f(p_i^+))y_i + e^{-af(p_i^+)}\min(p_i,\bar{p})(1-y_i)]}{\sum_{i\in I} v_i(e^d y_i + e^{-af(p_i^+)}(1-y_i))} \qquad (17)
\end{aligned}$$

The last simplification is obtained by injecting $y_i^2 = y_i$, i.e., $y_i(1-y_i) = 0 \quad \forall i \in I$.

3.2.2. Formulation:

The optimization problem for Model 2 can be formulated as follows.

Theorem 3.3 (Heterogeneous Model)**.** *Model* $(M2)$ *is a linear-fractional programming problem with binary variables:*

$$\begin{aligned}
(M2) \ \min_y \quad & V \frac{\sum_{i\in I}((v_i e^d y_i + v_i e^{-af(p_i^+)}(1-y_i))\min(p_i,\bar{p}) + f(p_i^+)v_i e^d y_i)}{\sum_{i\in I}(v_i e^d y_i + v_i e^{-af(p_i^+)}(1-y_i))} \\
s.t. \quad & \sum_{i\in I}(q_i - (1+\alpha)\bar{q})y_i \geq 0 \\
& \sum_{i\in I} \pi_i v_i y_i \geq \sum_{i\in I} \pi_i v_i - \mu V \\
& y_i \in \{0,1\} \ \forall i \in I \qquad (18)
\end{aligned}$$

Proof. We use Eq. (17) for the objective and Eqs. (7)-(8) of Model 1. □

In what follows, we assume that the trade-off parameters (μ, α) are selected by the decision maker such that the feasible set is not empty. Specifically, let Q^- be the set of providers with quality strictly below $(1+\alpha)\bar{q}$ and Q^+ the set of providers with quality at least equal to $(1+\alpha)\bar{q}$. Further, let $\tilde{\mathbf{y}}^*$ be an optimal solution of the following binary linear optimization problem on y_i, $i \in Q^-$ (we know an optimal solution exists because $\tilde{\mathbf{y}} = \mathbf{0}$ is feasible and the feasible set is bounded):

$$\begin{aligned}
\max_{\tilde{y}} \ & \sum_{i\in Q^-} \pi_i v_i \tilde{y}_i \\
\text{s.t.} \ & \sum_{i\in Q^-} ((1+\alpha)\bar{q} - q_i)\tilde{y}_i \leq \sum_{j\in Q^+} (q_j - (1+\alpha)\bar{q}) \\
& \tilde{y}_i \in \{0,1\} \ \forall i \in Q^-. \qquad (19)
\end{aligned}$$

Proposition 3.4 (Feasibility conditions). *It is necessary and sufficient for the feasible set of* $(M2)$ *to be non-empty to have*

$$\mu \geq \frac{1}{V}\left(\sum_{i\in I/Q^+}\pi_i v_i - \sum_{i\in Q^-}\pi_i v_i \tilde{y}_i^*\right)$$

Proof. For the feasible set of $(M2)$ to be non-empty, it is necessary and sufficient for the optimal objective of Problem (20) below to be at least $\sum_{i\in I}\pi_i v_i - \mu V$.

$$\begin{aligned} &\max_y \sum_{i\in I}\pi_i v_i y_i \\ &\text{s.t.} \sum_{i\in I}(q_i - (1+\alpha)\bar{q})y_i \geq 0 \\ &\quad y_i \in \{0,1\}\ \forall i \in I. \end{aligned} \tag{20}$$

It is always optimal to select the decision variables with non-negative weights in the constraint, leaving us to solve Problem (19). □

Remark: Problem (M2) is NP-hard since a simpler case where (i) the objective function is linear instead of fractional and (ii) quality constraints are removed (which can be done if all q_i are identical and $\alpha = 0$), is NP-hard itself. This simpler case can be written in a generic way as $\min \sum_i c_i y_i$ s.t. $\sum_i a_i y_i \geq b,\ y_i \in \{0,1\}$ which is equivalent to an (NP-hard) knapsack problem replacing binary variables y_i by $(1 - x_i)$, x_i binary.

3.2.3. Solving the Problem using Fractional Optimization:

We can rearrange $(M2)$ in the form of a linear-fractional problem with binary variables and a strictly positive denominator:

$$\begin{aligned} (F)\quad \min_{\mathbf{y}}\quad & z = \frac{c_0 + \mathbf{cy}}{g_0 + \mathbf{gy}} \\ s.t\quad & \mathbf{Ay} \geq \mathbf{b} \\ & \mathbf{y} \in \{0,1\} \end{aligned}$$

Based on Isbell and Marlow (1956), Dinkelbach (1967) developed a parametrization method to solve this. The key idea is to use a linear parametrized problem $(F_{par}(\zeta))$ for a given $\zeta \in \mathbb{R}$:

$$\begin{aligned} (F_{par}(\zeta))\quad \min_{y}\quad & h(\zeta) = \mathbf{cy} - \zeta\mathbf{gy} \\ s.t\quad & \mathbf{Ay} \geq \mathbf{b} \\ & \mathbf{y} \in \{0,1\} \end{aligned}$$

Although Dinkelbach uses the objective function $h'(\zeta) = c_0 + \mathbf{cy} - \zeta(g_0 + \mathbf{gy})$, it is immediate to see that removing the constant term leaves the core problem unchanged. Dinkelbach proved that the optimal solution of the initial problem (F) is the same as the optimal solution of $F_{par}(\zeta^*)$ if and only if $h(\zeta^*) = g_0\zeta^* - c_0$. Then $z^* = \zeta^*$, and the goal is to find ζ such that $h(\zeta) - (g_0\zeta - c_0) = 0$.

Algorithm 1 Megiddo algorithm for a problem (F)

Initialize (ζ_l, ζ_u)
while $|g_0\zeta - c_0 - h^*(\zeta)| \geq \epsilon$ **do**
  $\zeta \leftarrow \frac{\zeta_l+\zeta_u}{2}$
  Solve $F_{par}(\zeta)$
  $y_\zeta \leftarrow$ optimal solution
  if $g_0\zeta - c_0 - h^*(\zeta) \leq -\epsilon$ **then**
    $\zeta_l \leftarrow \zeta$
  else
    if $g_0\zeta - c_0 - h^*(\zeta) \geq \epsilon$ **then**
      $\zeta_u \leftarrow \zeta$
    end if
  end if
end while
output y_ζ

Megiddo (1979) developed an exact polynomial algorithm (see Algorithm 1) based on a dichotomous search of ζ to solve (F), which we apply to $(M2)$. The parametrized problem is:

$$
\begin{aligned}
M2_{par}(\zeta) \quad \min_{y} \quad & h(\zeta) = \mathbf{cy} - \zeta\mathbf{gy} \\
& = \sum_{i\in I}[(e^d - e^{-af(p_i^+)})(V\min(p_i,\bar{p}) - \zeta) + Ve^d f(p_i^+)]v_i y_i \\
\text{s.t} \quad & \sum_{i\in I}(q_i - (1+\alpha)\bar{q})y_i \geq 0 \\
& \sum_{i\in I}\pi_i v_i y_i \geq \sum_{i\in I}\pi_i v_i - \mu V \\
& y_i \in \{0,1\} \ \forall i \in I
\end{aligned}
$$

Initialization of $(\zeta_\mathbf{l}, \zeta_\mathbf{u})$ *for problem* (P)*:.* Since $M2_{par}(\zeta)$ converges to the optimal solution of $(M2)$, the initial search interval for ζ can be a lower bound and an upper bound of z_2^*.

Proposition 3.5. *Lower and upper bounds of* z_2^* *are* $(V\min_{i\in I}(p_i), V\frac{\sum_{i\in I} v_i e^d(\min(p_i,\bar{p}) + f(p_i^+))}{\sum_{i\in I} v_i e^{-af(p_i^+)}})$.

Proof. We are looking for an upper bound of:

$$
z_2 = V\frac{\sum_{i\in I} v_i e^{-af(p_i^+)(1-y_i)+dy_i}(\min(p_i,\bar{p}) + f(p_i^+)y_i)}{\sum_{i\in I} v_i e^{-af(p_i^+)(1-y_i)+dy_i}}
$$

We notice that $\forall i \in I, y_i \in \{0,1\}$: $v_i e^{-af(p_i^+)(1-y_i)+dy_i} \leq v_i e^d$ and $\min(p_i,\bar{p}) + f(p_i^+)y_i \leq \min(p_i,\bar{p}) + f(p_i^+)$, and $v_i e^{-af(p_i^+)(1-y_i)+dy_i} \geq v_i e^{-af(p_i^+)}$. Then $V\sum_{i\in I}\frac{v_i e^d(\min(p_i,\bar{p}) + f(p_i^+))}{\sum_{i\in I} v_i e^{-af(p_i^+)}} \geq z_2$. The lower bound is immediate to compute by remarking that: $(\min(p_i,\bar{p}) + f(p_i^+)y_i) \geq \min_{i\in I}(p_i) \quad \forall i \in I$. $\square$

3.2.4. Insights into the Optimal Strategy:

We partition the set of providers I into four subsets:

Low price, low quality $I_{LL} = \{i \in I \mid p_i \leq \bar{p}, q_i \leq (1+\alpha)\bar{q}\}$,
Low price, high quality $I_{LH} = \{i \in I \mid p_i \leq \bar{p}, q_i > (1+\alpha)\bar{q}\}$,
High price, low quality $I_{HL} = \{i \in I \mid p_i > \bar{p}, q_i \leq (1+\alpha)\bar{q}\}$,
High price, high quality $I_{HH} = \{i \in I \mid p_i > \bar{p}, q_i > (1+\alpha)\bar{q}\}$.

Note that the concepts of low/high price and low/high quality are subjective in the sense that the respective thresholds to define low and high are the reference price $\bar{p}$ and the improved average quality $(1+\alpha)\bar{q}$. It is clear that $I_{LL} \cup I_{LH} \cup I_{HL} \cup I_{HH} = I$.

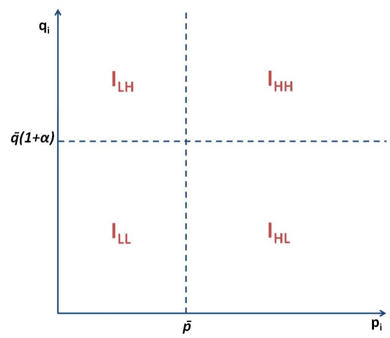


Figure 2: Partition of the set of providers I by price and quality

Theorem 3.6. *Low-price, high-quality providers ($i \in I_{LH}$) are exempted from RP when their price falls below a common threshold. Further, if there exists a high-price, high-quality provider ($i \in I_{HH}$) exempted from RP, then all providers in I_{LH} are exempted from RP.*

Proof. Assume we have found the optimal ζ^* for which to solve $M2_{par}(\zeta)$. Then for $i \in Q^+$ (high-quality providers), setting $y_i = 1$ never decreases the slack to the constraints and it is optimal to set $y_i = 1$ when the coefficient in front of y_i in the objective is non-positive. For $i \in I_{LH}$, this coefficient is $v_i(e^d - 1)(Vp_i - \zeta^*)$, i.e., it is optimal to set $y_i = 1$ when $p_i \leq \zeta^*/V$. For $i \in I_{HH}$, this coefficient is $v_i[(e^d - e^{-af(p_i^+)})(V\bar{p} - \zeta^*) + Ve^d f(p_i^+)]$, which is non-positive if $\bar{p} < \zeta^*/V$. Since $\bar{p} \geq p_i$ for all $i \in I_{LH}$, we do then immediately have that $p_i \leq \zeta^*/V$ for all $i \in I_{LH}$. □

Below, we call the *knapsack-motivated heuristic* the well-known heuristic for the knapsack problem that ranks items by the ratio of the objective to the constraint weight (Dantzig, 1957).

Theorem 3.7 (Heuristics for Extreme Cases)**.** *(i) Without the quality constraint, the solution of the knapsack-motivated heuristic for the provider selection strategy does not depend on initial patient volumes. (It is indirectly affected by quality through the satisfaction probabilities π_i)*
(ii) Without the satisfaction constraint, the optimal strategy for $i \in Q^+$ depends neither on q_i nor on v_i. The solution of the knapsack-motivated heuristic for $i \in Q^-$ does depend on q_i and v_i.

Proof. (i) Without the quality constraint, or for a given quality level, the ratio of the objective to the weight is independent of v_i (the v_i in the numerator and denominator cancel out.)
(ii) For $i \in Q^+$, the quality constraint is further slackened when $y_i = 1$. The decision maker will set y_i to 1 when the coefficient in the objective is non-positive. The sign of that coefficient is the sign of $(e^d - e^{-af(p_i^+)})(V \min(p_i, \bar{p}) - \zeta) + Ve^d f(p_i^+)$, which depends neither on quality nor on volume. For $i \in Q^-$, the heuristic uses $v_i[(e^d - e^{-af(p_i^+)})(V \min(p_i, \bar{p}) - \zeta) + Ve^d f(p_i^+)]/[(1+\alpha)\bar{q} - q_i]$. □

Remark: This suggests that volume most affects the strategy for low-quality providers when the satisfaction constraint has enough slack. Decreased slack in the satisfaction constraint drives the heterogeneous selection of high-quality providers (some are selected but not others).

Finally, let us investigate the ideal strategy, i.e., the provider selection/exemption strategy the decision maker would implement if there were no quality or satisfaction constraints. This is the strategy the problem converges toward as the quality and satisfaction constraints are relaxed.

Theorem 3.8 (Ideal strategy)**.** *(i) In the case where the constraints are relaxed, a provider who charges less than the reference price will be exempted if his price is below a certain threshold. This threshold is the system-wide equivalent price at optimality, defined by the ratio of optimal cost objective to volume.*
(ii) If it is optimal to exempt a provider who charges more than the reference price, then we have $\zeta^* > V\bar{p}$ *and the ideal strategy is the following: first, to exempt providers who charge at most the RP. Then, to exempt providers who charge more than the RP but whose expected excess price post RP implementation* $f(p_i^+)$ *falls below a certain threshold. This threshold increases with* $\zeta^*/V - \bar{p}$*, which is interpreted as the excess over the RP of the system-wide equivalent price at optimality.*

Proof. The result follows from analyzing the sign of the coefficient $(e^d - e^{-af(p_i^+)})(V \min(p_i, \bar{p}) - \zeta) + Ve^d f(p_i^+)$, which is equal when $p_i < \bar{p}$ to $(e^d - 1)(Vp_i - \zeta)$, allowing us to conclude in (i) by considering the sign of ζ/V and when $p_i \geq \bar{p}$ to $(e^d - e^{-af(p_i^+)})(V\bar{p} - \zeta) + Ve^d f(p_i^+)$. It is optimal to exempt provider i if and only if the coefficient is non-positive. A necessary condition is $V\bar{p} - \zeta < 0$. We then compute the value of $f(p^+)$ for which

$$(e^d - e^{-af(p^+)})(V\bar{p} - \zeta) + Ve^d f(p^+) = 0. \tag{21}$$

Because the left-hand side is negative for $f(p^+) = 0$ and increases in $f(p^+)$ (the term $(\zeta - V\bar{p})a$ is much smaller than Ve^d for the values of a described in the calibration of Section 4.2), either the equality has a unique solution or we set the threshold to ∞. When $\zeta/V - \bar{p}$ increases, the solution of Eq. (21) has to increase so that the left-hand side of Eq. (21) remains equal to zero. □

The following corollary connects the ideal strategy with the real constrained problem.

Corollary 3.1. *If it is optimal to exempt a high-quality provider* $i \in Q^+$ *from RP in the ideal strategy, then it is also optimal to exempt him in the in-network selection model* $(M2)$.

Proof. A high-quality provider i was exempted from RP in the ideal strategy when the corresponding objective coefficient was non-positive in $(M2_{par}(\zeta^*))$. Setting y_i to 0 would make the objective no better and decrease the slack in the constraints. Hence, keeping y_i at 1 makes the problem no worse, and strictly better if the objective coefficient was strictly negative. □

Remark: If it is optimal to exempt a low-quality provider $i \in Q^-$ from RP in the ideal strategy, then whether it is also optimal to do so in the real-world strategy (model $(M2)$) depends on the slacks of the quality and satisfaction constraints and the weight of that provider in the objective, i.e., how negative the coefficient was in the ideal strategy. This represents the trade-offs the decision maker would make by having average quality of exempted providers slightly diminish while improving the objective (decreasing the cost) and the satisfaction constraint.

3.3. Advantages and Drawbacks of the two models

Modeling accuracy: Model 2 captures the impact of out-of-pocket payments on volumes and patients' choice of a provider. In Model 1, the only feature that influences the patients' choice of provider is the binary information as to whether a provider has been exempted from RP. The exemption is perceived as a "stamp" awarded by the payer to providers showing a fair price-to-quality ratio. In Model 2 however, the patients' choice of provider is also influenced by the out-of-pocket price. The wide range of these out-of-pocket prices leads to differentiate the variation rates of volumes among hospitals, which can better capture reality if these variation rates are correctly captured by the parameters of the choice model.

Assumptions/data required: The strongest advantage of Model 1 is a relatively low requirement regarding parameter estimation. When compared to Model 2, parameter β_1 is more straightforward to intuitively understand and evaluate than choice model parameters (a, d). The CalPERS study (Cowling, 2013) provides an empirical estimate of β_1. The difficulty to calibrate choice model parameters, whenever empirical data are lacking, naturally leads to consider a robust approach.

Computation time: Model 2 solves instances of size $n = 150$ within seconds (the average number of iterations in the Megiddo algorithm is thirty), when Model 1 might need several minutes.

Practicality: Although Model 2 is less intuitive partly because of the exponential form of the choice model and partly because of the set of parameters to explain, its structure is easier to understand because it only uses binary choice variables y_i as decision variables.

4. Numerical Illustration

4.1. Data Simulation

For our numerical experiments, we generated independent random vectors for the providers' prices, volumes, and quality:

- Quality scores q_i are drawn from the set $\{1, 2, 3, 4, 5\}$ with a distribution inspired by the Centers for Medicare and Medicaid Services (CMS) 5-star rating system. Details on the exact distribution of stars can be found in CMS Five-Star Quality Rating System Technical Users' Guide (2012).
- Volumes are drawn uniformly from the interval $[10, 100]$. This range is based on CalPERS pilot study (Cowling, 2013) for the minimum volume and on market-sizing estimation for the maximum volume (with the assumption that the provider with the larger volume has 10 times more operations than the provider with the smallest volume).
- Prices are generated according to a combination of two power laws, the parameters of which provide the best-fit match of the chart of prices available in Figure 1 (CALPers empirical study). The distribution of prices observed on this chart can be reproduced by fitting, to the left of a certain threshold at approximately 2/3 of the horizontal axis, a power law of the form $X_{max}x^{\alpha_1}$ (with $0 < \alpha_1 < 1$), and to the right of this threshold, an inverse power law of the form $X_{min}(1-x)^{\alpha_2}$ (with $-1 < \alpha_2 < 0$).

A practitioner willing to implement the model could easily acquire historical data for prices and volumes. Care quality is more difficult to measure or even to define: it could be argued that patient satisfaction or provider reputation should be taken into account in an aggregate measure accounting for patients' choice. For the sake of model clarity, we choose to base our work on a simple but reliable CMS 5-star rating system.

We show results with a sample size of $n = 150$ and a reference price of $\bar{p} = 30$. We set the quality improvement parameter to $\alpha = 0.25$. Dissatisfaction probabilities π_i are set between 0% and 15%, and increase with quality q_i and out-of-pocket price $(p_i - \bar{p})^+$, in that lexicographic order. Therefore, if his provider is not selected to be exempted from RP, the patient is assumed to be all the more so dissatisfied that first, the provider quality is high, and second, the patient knows he has to pay a large amount above the reference price. For example, the dissatisfaction probability related to a provider of quality 2 and twice the reference price is 4%, and it is 6% for a provider of quality 3 but whose price is below the reference price. The bound on the total dissatisfaction volume is set to $\mu = 2\%$. These are adjustable parameters that can vary upon the practical application. Finally based on CalPERS pilot study, we assume that f is a linear function: $f(x) = 0.4x$.

4.2. Model Calibration

Estimation of β_1 for Model 1: This parameter quantifies the homogeneous volume increase observed/anticipated at providers in the network. Based on the CalPERS studies (Cowling, 2013), we set $\beta_1 = 20\%$.

Estimation of a and d for Model 2: These parameters quantify the change in volume related to the change in out-of-pocket price, or to the exemption/inclusion of a provider in the RP scheme. Although a is connected to the price elasticities of healthcare demand, studied for example in the RAND Health Insurance Experiment (Manning et al., 1987) we cannot simply identify a with a price elasticity of healthcare demand because of the exponential nature of our model. We set a and d to 0.01 so that provider volumes do not increase or decrease by more than 50%. Note that this assumption is rather conservative, since according to the expert in CalPERS pilot study a few providers with a low volume have a much larger relative increase or decrease.

4.3. Results

This section documents the practical insights decision makers can gain from our framework.

4.3.1. Representation of selected providers on a (p, q) plane:

On Figure 3, data points in a dark color represent providers who are exempted from RP, and in a light color represent those who are subject to RP. The size of data points represents volumes.

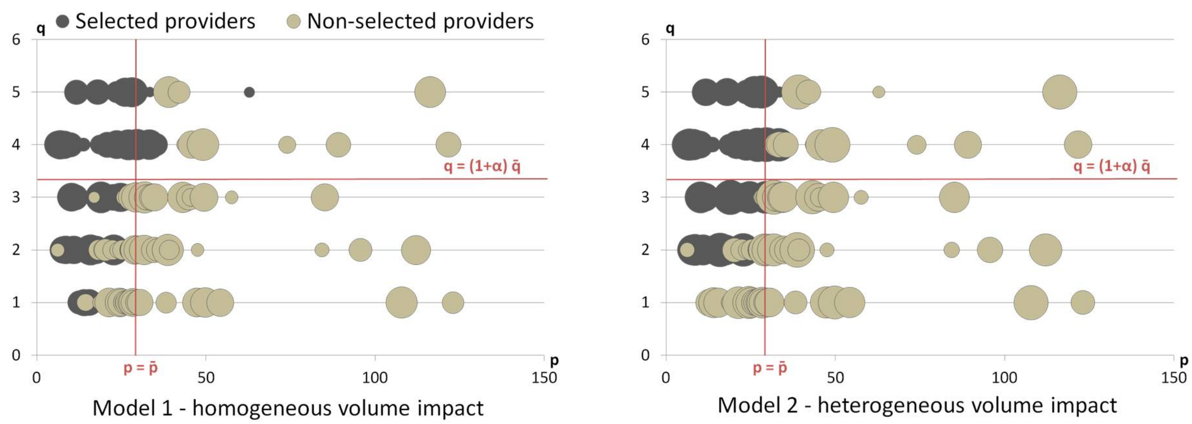


Figure 3: Selected providers on a (p, q) plane (default parameters)

We use the partition in four quadrants from Figure 2 to analyze the results. All providers of I_{LH} (low price, high quality) are exempted from RP both for the homogeneous and heterogeneous model, which matches our theoretical results, and no providers from I_{HL} (high price, low quality). The behavior in I_{HH} (high price, high quality) is also similar in the two models, with a very small number of providers of highest quality and price close to the RP exempted. We also observe notable differences between the two models by quality level: the homogeneous model exempts more extreme providers on the graph (providers with very low quality and price, or best quality and high price); as a result of the overall trade-off, less average quality providers are selected.

4.3.2. Influence of the Constraint Parameters on Model 1:

We analyse the impact on the network selection when the constraint parameters (α, μ) vary. This section is key from an application perspective, since these parameters are set by the decision maker when implementing the model. As mentioned in introduction, they can be regarded as levers for strategically shaping the network of providers. We use a pair of low/high values for α, combined with a pair of low/high values for μ. We show the results for Model 1 only for the sake of brevity, since the analysis for Model 2 is very similar.

First, the impact of the quality constraint is clearly seen on the top two quadrants of Figure 4, where α takes values of 15% and 35% instead of its default value of 25% and the satisfaction parameter μ is kept constant. For a low α (left panel), the frontier between exempted and non-exempted providers is almost vertical, and almost blended with the reference price limit. In other words, the optimal solution is very close to the ideal strategy identified in Theorem 3.2. When the payer makes little requirement about the quality of the exempted providers, he has little interest in selecting "expensive" high-quality providers, who will negatively impact the cost. In contrast, for a high α (i.e., a more stringent requirement on average quality to justify RP exemption), the frontier between exempted and non-exempted providers shifts to a diagonal: in order to satisfy the constraint, the payer selects more high-quality providers and less low-quality providers, only exempting a few lower-quality providers so that the satisfaction constraint is met.

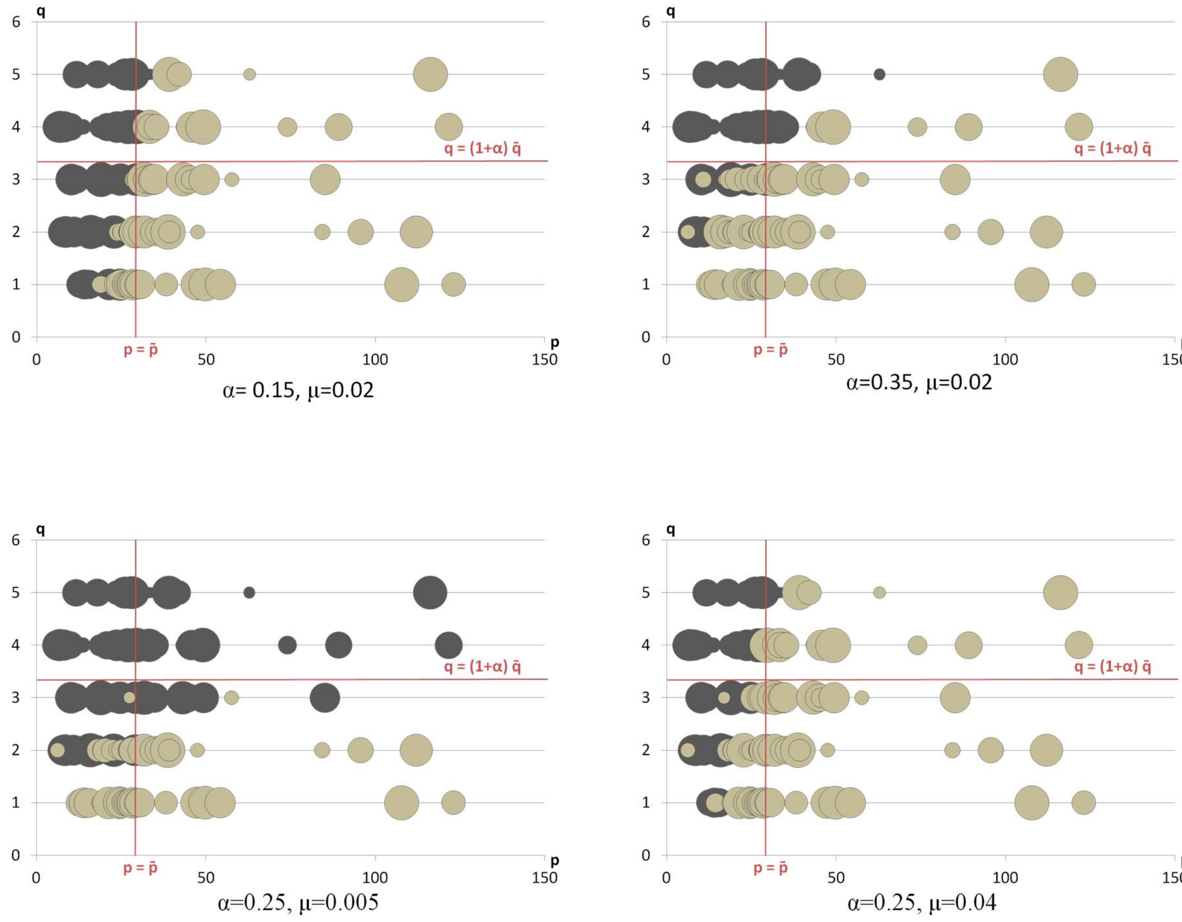


Figure 4: Influence of constraint parameters (α, μ) on Model 1 selection

Figure 4 also shows the impact of the satisfaction constraint. On the bottom-left graph, the total allowed proportion of dissatisfied patients is set to 0.5% instead of 2%. All top-quality providers are then exempted from RP, even the most expensive ones, because top-quality providers have the highest dissatisfaction probabilities. Below the quality limit of $(1+\alpha)\bar{q}$, providers are selected according to a combination of high quality, low price and high volume. On the bottom-right part of Figure 4 where the total allowed proportion of dissatisfied patients is relaxed to 4%, almost no provider with a price above the RP is exempted.

4.3.3. Influence of the Choice Parameters on Model 2:

A sensitivity analysis on choice parameters is available upon request (reference omitted for blind peer review). However, a more structured approach is to introduce uncertainty via robust optimization, which we conduct in the next section for both models.

5. Uncertainty on Demand Parameters

5.1. Robust Homogeneous Model

Regarding Model 1, we have assumed so far that all selected providers observe the same exogenous volume increase β_1 and all non-selected providers observe the same volume decrease β_2. In this section, we introduce the possibility of an uncertain deviation from these nominal values, for which we use robust optimization with interval uncertainty (as formulated in Ben-Tal and Nemirovski, 2002), where $\beta_1 \in [\underline{\beta}_1, \overline{\beta}_1]$. The robust problem can be formulated as the following non-linear problem, where β_2 has been replaced by its expression in constraint (4):

$$
\begin{aligned}
(M1Ra) \min_{y} \quad & [\max_{\beta_1} z(\beta_1) = \sum_{i \in I} v_i [\min(p_i, \bar{p})(1 - \beta_1 \frac{\sum_{k \in I} v_k y_k}{\sum_{k \in I} v_k (1 - y_k)}) \\
& + (\min(p_i, \bar{p})(\beta_1 + \beta_1 \frac{\sum_{k \in I} v_k y_k}{\sum_{k \in I} v_k (1 - y_k)}) + f(p_i^+)(1 + \beta_1)) y_i] \\
s.t \quad & \beta_1 \in [\underline{\beta}_1, \overline{\beta}_1] \qquad] \\
& \sum_{i \in I} q_i y_i \geq (1 + \alpha) \bar{q} \sum_{i \in I} y_i && (22) \\
& \sum_{i \in I} \pi_i v_i (1 - y_i) \leq \mu V && (23) \\
& y_i \in \{0, 1\} \forall i \in I && (24)
\end{aligned}
$$

Theorem 5.1 (Reformulation of Robust Homogeneous Model)**.** *Problem* $(M1Ra)$ *is equivalent to the fol-*

lowing linear problem:

$$
\begin{aligned}
(M1R)\ \min_{y} \quad & z \\
s.t \quad & z \geq \sum_{i\in I} v_i[\min(p_i,\bar{p})(1-\underline{w}_i) + y_i(\min(p_i,\bar{p})\underline{\beta}_1 + f(p_i^+)(1+\underline{\beta}_1))] \\
& z \geq \sum_{i\in I} v_i[\min(p_i,\bar{p})(1-\overline{w}_i) + y_i(\min(p_i,\bar{p})\overline{\beta}_1 + f(p_i^+)(1+\overline{\beta}_1))] \\
& \sum_{i\in I} q_i y_i \geq (1+\alpha)\bar{q}\sum_{i\in I} y_i \\
& \sum_{i\in I} \pi_i v_i (1-y_i) \leq \mu V \\
& \sum_{i\in I} v_i(\underline{\beta}_1 y_i - \underline{w}_i) = 0 \\
& \underline{\beta}_2 - y_i \leq \underline{w}_i \leq \underline{\beta}_2 && \forall i \in I \\
& \underline{w}_i \leq 1 - y_i && \forall i \in I \\
& \sum_{i\in I} v_i(\overline{\beta}_1 y_i - \overline{w}_i) = 0 \\
& \overline{\beta}_2 - y_i \leq \overline{w}_i \leq \overline{\beta}_2 && \forall i \in I \\
& \overline{w}_i \leq 1 - y_i && \forall i \in I \\
& y_i \in \{0,1\}, \underline{w}_i, \overline{w}_i, \underline{\beta}_2, \overline{\beta}_2 \in [0,1] && \forall i \in I
\end{aligned}
$$

Proof. We transform problem $(M1Ra)$ into a tractable problem in two steps. In the subproblem in brackets, the objective $z(\beta_1)$ is linear in β_1, although the sign of the coefficient is not constant (depends on vector $\boldsymbol{y}$). The maximum in the subproblem is thus reached at one of the two extreme points of the uncertainty interval. With $\underline{\beta}_2 = \underline{\beta}_1 \frac{\sum_{k\in I} v_k y_k}{\sum_{k\in I} v_k(1-y_k)}$ and $\overline{\beta}_2 = \overline{\beta}_1 \frac{\sum_{k\in I} v_k y_k}{\sum_{k\in I} v_k(1-y_k)}$, we can then transform the subproblem in brackets into:

$$
\begin{aligned}
\min_{y,z} \quad & z \\
s.t \quad & z \geq \sum_{i\in I} v_i[\min(p_i,\bar{p})(1-\underline{\beta}_2) + (\min(p_i,\bar{p})(\underline{\beta}_1 + \underline{\beta}_2) + f(p_i^+)(1+\underline{\beta}_1))y_i] \\
& z \geq \sum_{i\in I} v_i[\min(p_i,\bar{p})(1-\overline{\beta}_2) + (\min(p_i,\bar{p})(\overline{\beta}_1 + \overline{\beta}_2) + f(p_i^+)(1+\overline{\beta}_1))y_i] \\
& (22)-(24), \quad z \in \mathbb{R}
\end{aligned}
$$

Then we follow the same linearization method as in Theorem 3.1 with variables $\underline{w}_i$ and $\overline{w}_i$. □

5.2. Robust Heterogeneous Model

The main difficulty for practitioners remains to estimate the choice parameters, even after implementation of a pilot. We develop in this section an optimization model that incorporates uncertainty in the logit parameters of Model 2. We combine the modeling of Bertsimas and Sim (2004), where an uncertainty budget is allocated to all uncertain parameters, with the results in Schaible (1976).

We start from the deterministic problem $(M2)$ and introduce interval uncertainty on (a, d). We assume a linear uncertainty structure for the parameters that affect our formulation. Parameters are allowed to vary differently for each provider. Our goal is to obtain relatively simple, tractable robust formulations to see how the binary decision vector $\mathbf{y}$ changes in presence of uncertainty. We want to immunize against the worst case scenario when uncertain parameters belong to the following uncertainty set:

$$\begin{aligned}
&e^{d_i} = e^{d^0} + \phi_i^+(e^{\overline{d}} - e^{d^0}) - \phi_i^-(e^{d^0} - e^{\underline{d}})\\
&e^{-a_i f(p_i^+)} = e^{-a^0 f(p_i^+)} + \psi_i^+(e^{-\overline{a} f(p_i^+)} - e^{-a^0 f(p_i^+)}) - \psi_i^-(e^{-a^0 f(p_i^+)} - e^{-\underline{a} f(p_i^+)})\\
&\sum_i (\phi_i^- + \phi_i^+) \leq \Gamma_d, \quad \sum_i (\psi_i^- + \psi_i^+) \leq \Gamma_a, \quad \phi_i^-, \phi_i^+, \psi_i^-, \psi_i^+ \in [0,1] \forall i \in I
\end{aligned} \tag{25}$$

Γ_d and Γ_a are uncertainty budget parameters, taking values in $[0, n]$.

Let $d^- = (e^{d^0} - e^{\underline{d}}), d^+ = (e^{\overline{d}} - e^{d^0}), a_i^- = (e^{-a^0 f(p_i^+)} - e^{-\underline{a} f(p_i^+)}), a_i^+ = (e^{-\overline{a} f(p_i^+)} - e^{-a^0 f(p_i^+)})$.
The robust problem can be formulated as:

$$(M2Ra) \quad \begin{aligned}
\min_y \quad [\max_{\phi^-, \phi^+, \psi^-, \psi^+} \quad & z = \\
V \frac{\sum_{i \in I} \{ e^{d^0} + \phi_i^+ d^+ - \phi_i^- d^-) v_i y_i [\min(p_i, \bar{p}) + f(p_i^+)] + (e^{-a^0 f(p_i^+)} + \psi_i^+ a_i^+ - \psi_i^- a_i^-) v_i (1 - y_i) \min(p_i, \bar{p}) \}}{\sum_{i \in I} \{ (e^{d^0} + \phi_i^+ d^+ - \phi_i^- d^-) v_i y_i + (e^{-a^0 f(p_i^+)} + \psi_i^+ a_i^+ - \psi_i^i a_i^-) v_i (1 - y_i) \}} & \\
s.t. \quad & (25)] \\
s.t \quad & (22) - (24)
\end{aligned}$$

Problem $(M2Ra)$ can be reformulated in a tractable manner, as indicated in Theorem 5.2. We use the following notations:

$$\alpha = V \sum_{i \in I} (e^{d^0} v_i y_i (\min(p_i, \bar{p}) + f(p_i^+)) + e^{-a^0 f(p_i^+)} v_i (1 - y_i) \min(p_i, \bar{p}))$$

$$\beta = \sum_{i \in I} (e^{d^0} v_i y_i + e^{-a^0 f(p_i^+)} v_i (1 - y_i))$$

$$c_{[4n,1]} = V \begin{bmatrix} d^+ v_i y_i (\min(p_i, \bar{p}) + f(p_i^+)) \\ \cdots \\ -d^- v_i y_i (\min(p_i, \bar{p}) + f(p_i^+)) \\ \cdots \\ a_i^+ v_i (1 - y_i) \min(p_i, \bar{p}) \\ \cdots \\ -a_i^- v_i (1 - y_i) \min(p_i, \bar{p}) \\ \cdots \end{bmatrix} d_{[4n,1]} = \begin{bmatrix} d^+ v_i y_i \\ \cdots \\ -d^- v_i y_i \\ \cdots \\ a_i^+ v_i (1 - y_i) \\ \cdots \\ -a_i^- v_i (1 - y_i) \\ \cdots \end{bmatrix} A_{[4n+2,4n]} = \begin{bmatrix} 1 \cdots 1 | 0 \cdots 0 \\ 0 \cdots 0 | 1 \cdots 1 \\ \hline 1 \cdots\cdots\cdots 0 \\ \vdots \\ 0 \cdots\cdots\cdots 1 \end{bmatrix}$$

$$b^T = (\Gamma_d, \Gamma_a, 1 ... 1), b \in \mathbb{R}^{4n+2}$$

Theorem 5.2 (Reformulation of Robust Heterogeneous Model). *Problem* $(M2Ra)$ *is equivalent to the following nonlinear problem:*

$$
\begin{aligned}
(M2R)\ \min_{y,u,\lambda}\ & \lambda \\
s.t\quad & u_1 + u_{i+2} + \lambda d^+ v_i y_i \geq V\, d^+ v_i y_i (\min(p_i,\bar{p}) + f(p_i^+)) \quad \forall i \\
& u_1 + u_{n+i+2} - \lambda d^- v_i y_i \geq -V\, d^- v_i y_i (\min(p_i,\bar{p}) + f(p_i^+)) \quad \forall i \\
& u_2 + u_{2n+i+2} + \lambda a_i^+ v_i (1-y_i) \geq V\, a_i^+ v_i (1-y_i) \min(p_i,\bar{p}) \quad \forall i \\
& u_2 + u_{3n+i+2} - \lambda a_i^- v_i (1-y_i) \geq -V\, a_i^- v_i (1-y_i) \min(p_i,\bar{p}) \quad \forall i \\
& -\Gamma_d u_1 - \Gamma_a u_2 - \sum_{j=3}^{4n+2} u_j + \lambda \sum_{i\in I} (e^{d^0} v_i y_i + e^{-a^0 f(p_i^+)} v_i (1-y_i)) \geq \\
& \qquad V \sum_{i\in I} (e^{d^0} v_i y_i (\min(p_i,\bar{p}) + f(p_i^+)) + e^{-a^0 f(p_i^+)} v_i (1-y_i) \min(p_i,\bar{p})) \\
& (22)-(24), u \geq 0, \lambda \in \mathbb{R}
\end{aligned}
$$

Proof. First, we linearize the subproblem in brackets, using Charnes and Cooper (1962):

$$\max \quad \frac{c^T x + \alpha}{d^T x + \beta} \quad s.t \quad Ax \leq b, \quad d^T x + \beta > 0, \quad x \geq 0$$

becomes

$$\max \quad c^T y + \alpha t \quad s.t \quad Ay \leq bt, \quad d^T y + \beta t = 1, \quad (x,t) \geq 0$$

We subsequently take the dual of this linearized problem as in Schaible (1976) and get:

$$
\begin{aligned}
\min \quad & \lambda \\
s.t \quad & A^T u + \lambda d \geq c \\
& -b^T u + \lambda \beta \geq \alpha \\
& u \geq 0, \lambda \in \mathbb{R}
\end{aligned}
$$

Compared to Schaible (1976), our problem has $x \geq 0$ as an additional constraint. As a consequence in the dual formulation of the original paper, the first equality constraint becomes an inequality. We apply this to the subproblem, and get the full problem $(M2R)$. □

We solve Problem $(M2R)$ iteratively as a series of MIP parameterized by λ: λ is searched by dichotomy, starting at the objective value of the deterministic problem.

5.3. Numerical Implementation and Analysis

5.3.1. Models Calibration:

We implement both robust models using the data generated in Section 4. We run the robust homogeneous model for different values of the uncertainty interval $[\underline{\beta_1}, \overline{\beta_1}]$: $[18\%, 22\%]$, $[15\%, 25\%]$, $[10\%, 30\%]$. In doing so, we consider that the decision maker has very little information about the volume increase β_1 at in-network providers: no probability distribution is available, and the only estimate that it is possible to take is that this parameter lies in a certain interval.

Regarding the heterogeneous model, we vary both intervals and uncertainty budgets:

- The uncertainty intervals for choice models parameters (a) and (d) are respectively $[\frac{\bar{a}}{s}, \bar{a}s]$ and $[\frac{\bar{d}}{s}, \bar{d}s]$, where $\bar{a}$ and $\bar{d}$ are the nominal parameters of the deterministic model (see Section 4) and s is a range coefficient taking values in set $\{5, 10, 100\}$. This method enables us to test the robustness of the model in intervals of different width.

- Uncertainty budgets Γ_d and Γ_a take values in set $\{0, 30, 60\}$. These numbers can be interpretated as how many providers will see their demand parameters deviate from the nominal value.

5.3.2. Results.

Impact on the objective:. Tables 3 and 4 show the percentage of increase of the objective function when we introduce uncertainty on parameters. Recall that Bertsimas and Sim (2004) found that worst-case protection against controlled uncertainty (with Γ budget) only provides a low loss of value empirically. This is confirmed by Table 4. The tables also confirm the intuition that the cost gets worse as the uncertainty increases, since the robust model accounts for the worst-case scenario in a given interval. More importantly, they give an idea of the cost impact of uncertain parameters for decision makers using the model. For example, if the decision maker has very little information about the volume increase at in-network providers β_1 and estimates this proportion to be between 10% and 30% instead of exactly 20%, the estimated cost of the selected network will increase by 1.5%. This cost sensitivity can be taken into account by the decision maker when calibrating the model.

Table 3: Increase in % of the objective function with uncertainty - homogeneous model

$[\underline{\beta_1}, \overline{\beta_1}]$	18-22%	15-25%	10-30%
	0.3%	0.8%	1.5%

Table 4: Increase in % of the objective function with uncertainty - heterogeneous model

s \ (Γ_d, Γ_a)	(30,0)	(0,30)	(60,0)	(0,60)	(30,30)	(60,60)
5	1.9%	0.2%	2.0%	0.2%	2.0%	2.2%
10	2.2%	0.4%	2.3%	0.4%	2.5%	2.7%
100	2.5%	4.6%	2.7%	4.7%	6.3%	6.4%

Impact on the providers tiers:. We also present the results in terms of network selection on the (p, q) plane. On Figure 5, we highlight in a darker color providers that change tiers when we introduce uncertainty. It is noteworthy to mention that even with very wide uncertainty intervals for parameters, only 11 providers out of 150 are subject to change for the homogeneous model, and 7 providers for the heterogeneous model. They are located, understandably, on the diagonal frontier between the two tiers. In other words, the optimal decision of whether to include them in the network or not is straightforward regarding most providers; however for a few providers, the optimal decision is not the same for different values of the parameters. If needed be, the decision maker can look at additional criteria (historical relationship or contracting with a provider, requirement on the total size of the network...) in order to decide whether to include these remaining providers or not.

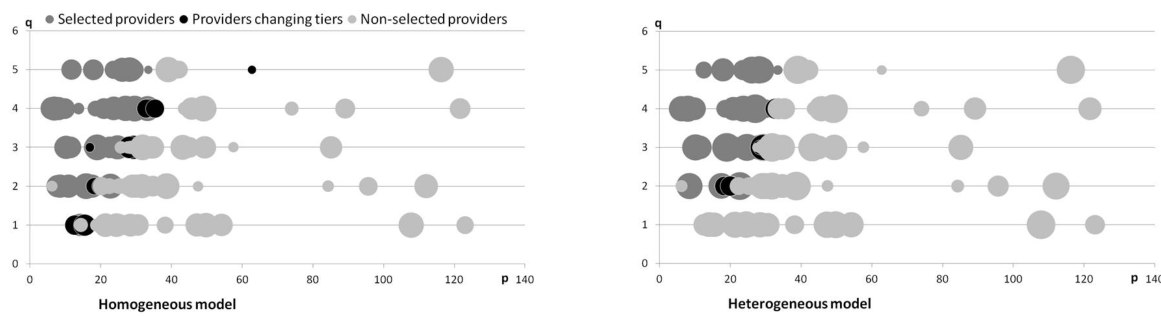


Figure 5: Impact of parameter uncertainty on provider tiers

6. Conclusion

This paper develops a structured framework for practitioners willing to pursue the experiment of reference pricing for healthcare and wish to incorporate several criteria in the selection of providers. It presents two optimization models: the homogeneous model roughly captures volume increase at in-network providers and can be used when little data about patients' preferences is available, whereas the heterogeneous model captures flows of patients based on their preferences. One important theoretical result, in Theorem 3.8, establishes a threshold policy to determine which providers should be exempted from reference pricing when the external constraints are relaxed. Section 4.3.2 shows how, starting from that, the constraints (specifically the constraint parameters) can shape the solution.

The model we develop is of potential interest outside the healthcare context. It could be generalized to a network selection problem where the demand and prices depend on the network itself. We also identify several research avenues relative to this specific problem. First, more empirical research can be performed on how the different criteria weight in patients' choice of a provider. From a modeling point of view, it could also be interesting to incorporate the reference price itself as a decision variable, possibly in a two-stage model.

Rising healthcare prices are now a major concern in the USA, and this kind of framework will prove to be of primary interest for all stakeholders. We should emphasize that a value-oriented payment system such as this adapted version of RP cannot function without collaboration between payers and providers, as well as an improved transparency of price and quality.

Bibliography

Andritsos, D. A., Tang, C. S., 2014. Introducing competition in healthcare services: The role of private care and increased patient mobility. European Journal of Operational Research 234 (3), 898–909.

Ben-Tal, A., Nemirovski, A., 1999. Robust solutions of uncertain linear programs. Operations Research Letters 25 (1), 1–13.

Ben-Tal, A., Nemirovski, A., 2002. Robust optimization–methodology and applications. Mathematical Programming 92 (3), 453–480.

Bertsimas, D., Sim, M., 2004. The price of robustness. Operations Research 52 (1), 35–53.

Brailsford, S., Vissers, J., 2011. Or in healthcare: A european perspective. European journal of Operational research 212 (2), 223–234.

Brekke, K. R., Königbauer, I., Straume, O. R., 2007. Reference pricing of pharmaceuticals. Journal of Health Economics 26 (3), 613–642.

Centers for Medicare and Medicaid services, 2012. Design for nursing home compare five-star quality rating system: Technical users guide. Tech. rep.

Charnes, A., Cooper, W. W., 1962. Programming with linear fractional functionals. Naval Research Logistics Quarterly 9 (3-4), 181–186.

Cowling, D., 2013. Calpers reference pricing program for hip or knee replacement, 2013.

Dantzig, G. B., 1957. Discrete-variable extremum problems. Operations Research 5 (2), 266–288.

Danzon, P. M., Ketcham, J. D., 2003. Reference pricing of pharmaceuticals for Medicare: evidence from Germany, the Netherlands and New Zealand. Vol. 10007. National Bureau of Economic Research.

Dellino, G., Kleijnen, J. P., Meloni, C., 2010. Robust optimization in simulation: Taguchi and response surface methodology. International Journal of Production Economics 125 (1), 52–59.

Dellino, G., Kleijnen, J. P., Meloni, C., 2012. Robust optimization in simulation: Taguchi and krige combined. INFORMS Journal on Computing 24 (3), 471–484.

Dinkelbach, W., 1967. On nonlinear fractional programming. Management Science 13 (7), 492–498.

Emanuel, E., Tanden, N., Altman, S., Armstrong, S., Berwick, D., de Brantes, F., Calsyn, M., Chernew, M., Colmers, J., Cutler, D., et al., 2012. A systemic approach to containing health care spending. New England Journal of Medicine 367 (10), 949–954.

Gabrel, V., Murat, C., Thiele, A., 2014. Recent advances in robust optimization: An overview. European Journal of Operational Research 235 (3), 471–483.

Health Care Cost Institute, 2014. 2013 health care cost and utilization report. Tech. rep.

Heidenberger, K., Stummer, C., 1999. R & d project selection and resource allocation: a review of quantitative modelling approaches. International Journal of Management Reviews 1 (2), 197–224.

Ho, W., Xu, X., Dey, P. K., 2010. Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. European Journal of Operational Research 202 (1), 16–24.

Hsia, R. Y., Kothari, A. H., Srebotnjak, T., Maselli, J., 2012. Health care as a market good? Appendicitis as a case study. Archives of internal medicine 172 (10), 818–819.

Isbell, J., Marlow, W., 1956. Attrition games. Naval Research Logistics 3 (1-2), 71–94.

López-Casasnovas, G., Puig-Junoy, J., 2000. Review of the literature on reference pricing. Health policy 54 (2), 87–123.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., Leibowitz, A., 1987. Health insurance and the demand for medical care: evidence from a randomized experiment. The American economic review, 251–277.

McFadden, D., 2001. Economic choices. American Economic Review, 351–378.

Megiddo, N., 1979. Combinatorial optimization with rational objective functions. Mathematics of Operations Research 4 (4), 414–424.

Rais, A., Viana, A., 2011. Operations research in healthcare: a survey. International Transactions in Operational Research 18 (1), 1–31.

Reinhardt, U. E., 2013. The sleeper in health care payment reform.
URL `http://economix.blogs.nytimes.com/2013/08/02/the-sleeper-in-health-care-payment-reform`

Robinson, J. C., Brown, T., Whaley, C., 2015. Reference-based benefit design changes consumers choices and employers payments for ambulatory surgery. Health Affairs 34 (3), 415–422.

Robinson, J. C., Brown, T. T., 2013. Increases in consumer cost sharing redirect patient volumes and reduce hospital prices for orthopedic surgery. Health Affairs 32 (8), 1392–1397.

Rusmevichientong, P., Topaloglu, H., 2012. Robust assortment optimization in revenue management under the multinomial logit choice model. Operations Research 60 (4), 865–882.

Schaible, S., 1976. Fractional programming. i, duality. Management Science 22 (8), 858–867.

Soyster, A. L., 1973. Technical note - convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research 21 (5), 1154–1157.