Submitted to *Manufacturing & Service Operations Management*
manuscript (Please, provide the manuscript number!)

Optimal Selection of Health Care Providers

Dimitris Bertsimas, Jerry Kung
Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
{dbertsim@mit.edu, jkung@mit.edu}

Problem definition: Health care expenditures have been increasing at unsustainable rates for more than thirty years with no signs of abating. A significant proportion of increasing expenditures can be attributed to a lack of price transparency. We address the problem of reducing health care costs for self-insured employers with access to employees' claims data by identifying lower cost providers in an optimal way.

Academic/Practical Relevance: Our approach leads conservatively to a 10% reduction of health care costs in a real-world study involving a large private US employer. This underscores the potential of harnessing the power of large data sets and associated data-driven approaches towards alleviating one of the most pressing problems facing the United States today.

Methodology: We propose an approach based on mixed integer linear optimization that utilizes the data to generate decisions on how to optimally select providers to minimize cost, while respecting convenience for employees. We demonstrate that our approach is tractable for large-scale datasets and present the computational results on a real health care claims dataset.

Results: Our data-driven, optimization based approach leads to reduction of health care costs by 10% for a private employer in the southern United States with 14,000 dependents, by reassigning patients to different providers for a small number of procedures with a relatively small degree of employee inconvenience (an average (maximum) increase of health care related travel of 18km (40km)). These cost reductions are robust to changes in a variety of parameters.

Managerial implications: Our framework generates a list of low-cost providers capable of providing an assortment of procedures that adequately cover the existing patient population. This list could be used to suggest providers or inform potential partnerships for service contracts.

Key words: health care management, math programming, service operations

1. Introduction

For more than thirty years, the cost of health care in the United States (US) has been increasing at a rate far greater than in other industrialized countries around the world (Squires 2012). In 1980, total health care expenditures accounted for approximately 9% of the total gross domestic product (GDP) of the US. By 2009, it accounted for 17.4% of total GDP. In contrast, the country with the second-highest health spending as a percentage of GDP in 2009 was the Netherlands, at 12% of total GDP. The ballooning of health care costs in the US is likely to continue its upward trajectory; unless action is taken, spending on health care is projected to reach a quarter of total US GDP by 2037 (Squires 2012).

Per capita spending on health care (after adjusting for purchasing power parity) in the US was almost 80% higher than in Norway, the next highest country (Squires 2012). The greater expense of health care costs in the US has been attributed to higher prices than in other OECD countries. However, these higher prices have not translated to better quality. For instance, US hospitals are about average when measured by in-hospital mortality rates for heart attacks and stroke, and far below average for largely preventable deaths due to complications from asthma and diabetic amputations (Squires 2012). Additionally, the five-year survival rate for breast cancer, cervical cancer, and colorectal cancer is no higher in the US than in other OECD countries (Squires 2012).

One of the main drivers of higher prices in the US health care industry is the lack of transparency of how much a procedure costs (Beck 2014). The amount paid for a particular procedure is highly dependent on who is paying the bill and can differ, in some instances, by several orders of magnitude. For example, Medicare would reimburse $335 for an MRI at a hospital in Dearborn, Michigan, while some private insurance companies would pay up to $1,990 for the same procedure at the same hospital (Beck 2014). Some of these large pricing differentials are a result of the asymmetry of market power between private health insurance companies and larger regional hospitals.

Many experts also argue that prices vary as a result of different accounting methods that allow hospitals to recoup losses from more costly services, although this does not explain high costs for office vists (Beck 2014)

In addition to the opaque nature of price setting for medical procedures, many large-scale providers have implemented non-competitive contracting policies that lead to unjustified price differentials (Commonwealth 2011). These providers, often spread out across large geographical areas, force payers to negotiate on a single price for all locations, even if payers only wish to utilize one location. This leads to higher prices in locations where geographic differences do not warrant them and exacerbates the problem of large variations in price within a small area.

Many experts in the medical community have started calling for full transparency of pricing for medical procedures (Emanuel et al. 2012, Sinaiko and Rosenthal 2011). While some states such as California, Massachusetts, and New Jersey have started reporting basic summary statistics regarding charges for common procedures in hospitals in an effort to empower patients as they decide where to seek treatment, some have argued this transparency in and of itself does not lead to decreased costs (Sinaiko and Rosenthal 2011). In fact, since most patients are insured and thus immunized against much of the price variation above their deductibles, it is possible for such transparency to lead to increased costs on the part of the insurer, as the incentives for seeking cheaper care are severely reduced.

Many insurance companies have begun to develop reference prices for particular procedures; once reference prices are set, they will only cover costs up to the reference price. Patients are free to select their medical provider, but are responsible for any charges incurred above the reference price. The potential implications of data-driven pricing cannot be understated. Using reference pricing on hip and knee replacements alone, WellPoint, a large insurance company, was able to decrease prices in more expensive hospitals by up to one-third between 2008 and 2012 (Reinhardt 2013). With the advent of more plentiful electronic data, the time is ripe for employers to utilize data-driven decision-making to tackle the unsustainable explosion in health care costs in the US.

1.1. Our Contributions

In this paper, we propose a framework through which employers who self-insure can identify lower-cost medical providers that are convenient for their employees. Our framework utilizes historical electronic claims data to determine providers that are already charging reasonable prices and that are in close geographical proximity to the patient population. We discuss how to process general electronic claims datasets to extract the relevant information necessary to conduct this analysis.

Our main contributions are two-fold. First, we propose an approach based on mixed integer linear optimization that can be practically solved to optimality for large-scale datasets. We then demonstrate that the computational results returned by this approach can yield significant reductions in overall costs for medical procedures. We consider the sensitivity of our results to changes in a variety of our initial assumptions and observe that the cost savings are robust to these changes. Our modeling framework naturally lends itself to providing actionable decisions, thus bridging the gap between descriptive and prescriptive analytics. We also provide a discussion of how these prescriptions for optimally selecting medical providers can be realized in practice, so that these cost savings can be achieved.

1.2. Outline

In the next section, we discuss our dataset in more detail and provide some descriptive data analytics. We also formally define our problem of interest and detail our data cleaning procedures. In Section 3, we present a tractable mixed integer linear optimization approach that solves our problem of interest. In Section 4, we discuss our computational results and analyze the sensitivity of the results to different parameter settings. We also consider how the changes prescribed by our approach affect the prices paid for medical procedures, the number of procedures delivered by medical providers, and the distances that patients must travel. Section 5 delves into the policy implications of our work and addresses ways to implement changes in health insurance to achieve cost reductions for medical procedures. Section 6 concludes.

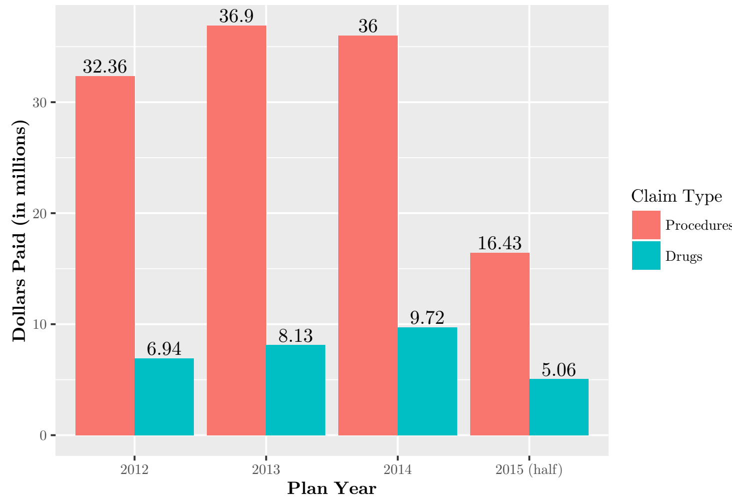


Figure 1 **Total costs paid out by the insurance plan for procedures and drugs over time. Note that our dataset ends in June 2015, so the visualization only captures half of Plan Year 2015.**

2. The Dataset

We obtained a dataset consisting of all health care claims paid out between 1 January 2011 and 1 July 2015 by a large self-insuring private employer based in the southern United States. Each row in the dataset contains information about an insurance claim that was filed for a procedure delivered to an employee or dependent, including a provider identification number, date of service, procedure name, and amount paid. We describe the data in more detail in Appendix A.

2.1. Descriptive Analytics

The dataset contains 1.55 million claims from more than 14,000 patients nationwide. These claims amounted to a total expenditure of more than \$151 million, of which approximately \$121 million were for medical procedures, and \$30 million were for prescription drugs. Approximately 30,000 providers and 4,792 medical procedures appear in the dataset. Figure 1 breaks down the costs of procedures and drugs by plan year. Note that costs from plan year 2015 are lower than preceding years because our dataset only includes claims from the first half of the year. Furthermore, note that the plan year of a claim may not correspond to the actual year in which the procedure occurred or prescription was filled, due to delays in billing.

The mean total cost (for procedures and prescription drugs) per patient in plan year 2012 was \$4032, increasing approximately 10% to \$4473 in plan year 2013. Although there was a slight

	Percentile of Patient				
Plan Year	20	40	60	80	Max
2012	$4,000	10,000	23,000	59,000	466,000
2013	5,000	13,000	30,000	88,000	521,000
2014	5,000	13,000	33,000	85,000	567,000
2015 (*)	3,000	8,000	20,000	45,000	283,000

Table 1 Quintiles of patient costs by plan year in dollars. Most patients incur relatively low cost, while a small percentage of the patients account for a large portion of total expenditures. (*) Data available only for the first half of 2015.

decrease in mean cost per patient in plan year 2014, halfway through plan year 2015, the mean total cost is $2415 per patient. This suggests that once all claims for 2015 are processed, there will be an increase in mean cost per patient by about 10% over that of plan year 2014. On the other hand, median cost per patient in plan year 2012 was $1011, with slight year-to-year decreases. Median cost per patient in plan year 2014 was $915. These summary statistics suggest that the distribution of patient costs is highly skewed.

We can take a deeper look into how costs per patient are distributed over the patient population by looking at the quintiles of the amount paid per patient over each plan year in Table 1.

Figure 2 shows the cumulative contribution to the total amount paid in each plan year after sorting all of the patients by their costs. For instance, we can see that in each of the plan years, the lowest 50% of all patients contribute only about 5% of the overall cost for that particular year. The top 10% of the most expensive patients contribute approximately 60% of all of the costs. The black line in the figure is a reference for perfectly equal distribution of costs over all of the patients. Overall, from this figure, we can see that the distribution of costs over the patients is highly unequal in any particular plan year and that this distribution remains roughly the same from year to year.

Both Table 1 and Figure 2 corroborate the findings in other studies Stanton (2005). In particular, we see that the top 5% of patients by amount paid accounted for approximately 55% of all costs in

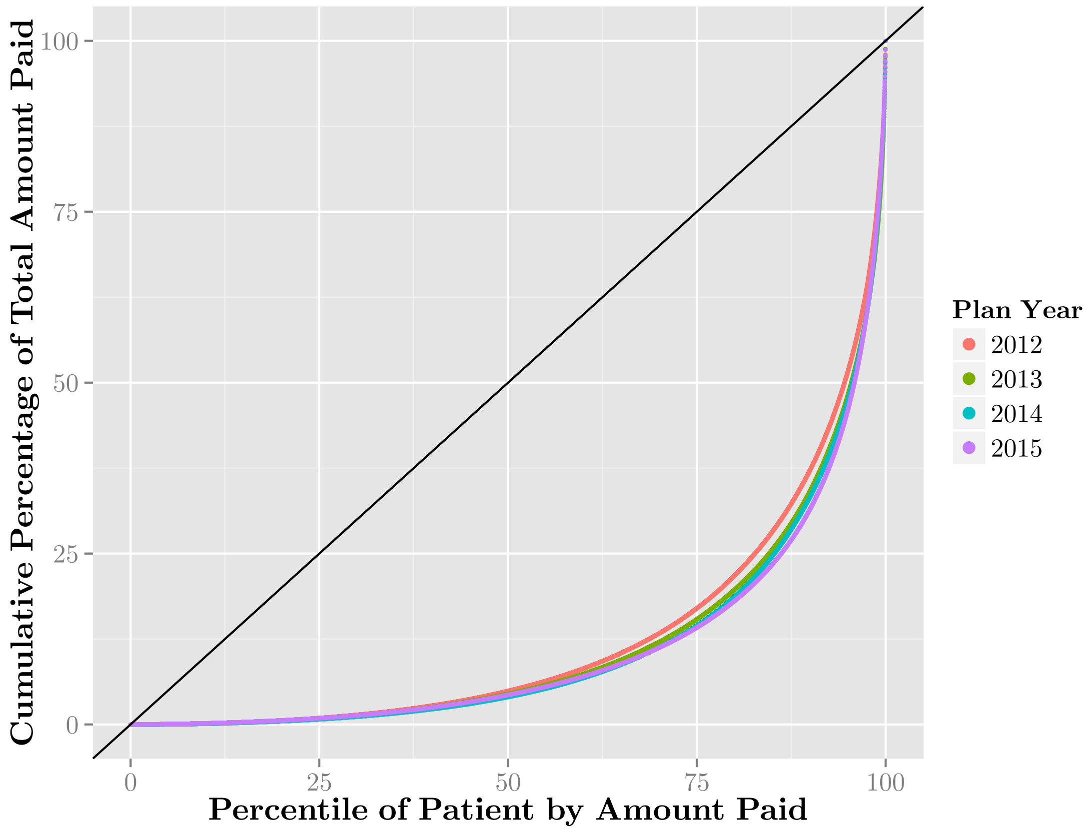


Figure 2 **Distribution of costs over patients for each of the four plan years. The black diagonal line represent perfect equality among all patients. The distribution of costs per patient is highly skewed, and we can observe that the top 5% of patients by amount paid account for a majority of healthcare expenditures. We also note that this distribution stays roughly the same from year to year.**

any given year. Conversely, the bottom 50% of patients by amount paid accounted for about 4% of all costs. This suggests that our dataset is reflective of the US population at large and indicates that our results could reasonably be extrapolated to a larger scale.

2.2. Problem of Interest

We aspire to identify a set of low-cost medical providers who can deliver a wide range of procedures within close geographical proximity of the patient population. Because our dataset includes information on how much was paid to medical providers for particular procedures in the past, we can identify providers who have costs that are consistently lower than other providers delivering the same procedure. In this way, we allow the data to inform our decisions as to which providers are already charging reasonable prices for the same procedures. Our data also gives us a good sense of the historical demand that these providers have satisfied from our patient population.

Given this data, our goal is to understand the magnitude of potential savings and to identify providers who would could help realize these reductions in cost. This collection of providers and their prices for procedures could serve as the starting point for further negotiations leading to even lower costs. We present the particular data cleaning procedures that are necessary to extract the inputs specific to our optimization model in Section 3.

2.3. Data Cleaning Procedures

The main aspiration of our work is to identify lower cost providers who are capable of delivering the same procedures while remaining in close geographical proximity to the patients' original providers. As a result, the two goals of our data cleaning process are to identify the actual locations where procedures occurred as well as the cost distribution for a specific provider-procedure pair.

Identifying Procedure Delivery Locations: Because we wish to understand price discrepancies for the same procedure within a particular geographic vicinity, our analysis relies on the accuracy of where particular procedures were provided.

Obtaining Costs for Provider-Procedure Pairs: Without properly understanding the true cost distribution for a specific provider-procedure pair, our approach might select providers based on prices that are too low, thus affecting the validity of our results.

Both of these data cleaning steps present several unique challenges and are crucial to the validity of our results. We describe them in more detail in Appendix B. Overall, our data cleaning procedures reduced the number of rows in our dataset from 1.5 million to about 520,000 and accounted for approximately $82.5M in costs.

3. Model Formulation

After completing the data preparation procedures from the previous section, each row in our dataset is structured as follows:

$$\texttt{claim} = (\texttt{PersonId, ProcedureName, ProviderName(s), LocationZIP, AmountPaid})$$

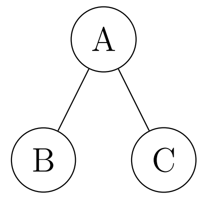
We think about our problem as follows. If a patient requires procedure k in a particular zip code z, which provider is within close proximity of z and can provide k at a low cost? For our analysis, we formulate a model whose solution represents perfect hindsight; that is, given the historical claims data that we have collected so far, we consider the optimal re-assignment of patients to nearby low-cost providers. As we discuss in the next section, most of the claims in the dataset are concentrated within a small geographical region. This allows us to disaggregate the optimization problem into smaller, separable subproblems, without sacrificing optimality.

3.1. Creating Service Regions

We restrict our attention to claims related to medical procedures. Additionally, we define a maximum acceptable travel distance that denotes a radius of how far we will are willing to look for lower-cost providers. For the purposes of our analysis, we have selected a maximum acceptable travel distance, which we denote by δ. We conduct sensitivity analysis on an appropriate selection of δ between 30 and 50 kilometers, a distance that patients can reasonably travel within 20-45 minutes.

Because we do not look for providers beyond a distance δ from the original service location, we can divide our problem into separable regions, where each region consists of zip codes that occur within our dataset. If we were to draw an edge between every pair of zip codes within δ from each other, a region would be equivalent to a connected component in the graph. To create our regions, we successively grow out each region until no additional zip codes within δ can be reached. This procedure produces a set of disjoint regions $r \in R$.

Given these regions, we can decompose the overall problem by solving a single optimization problem for each $r \in R$. By construction, for any $z, z' \in Z$ such that $z \in r, z' \in r', r \neq r'$, observe that any changes to the capacity or price of the provider at z' do not change our decision with respect to optimizing over region r. Thus, when identifying potential low-cost providers for an existing service location, we can solve our optimization problem separately over each of the individual regions, without sacrificing optimality.



	A	B	C
Capacity	10	20	20
Demand	10	10	10
Price	6	5	8

Figure 3 A small counter-example against greedy consideration of neighbors. Provider A is within δ of Providers B and C; Providers B and C are more than δ apart. However, a change in the price of Provider C from 8 to 4 will lead to changes in the utilization of Provider A.

In contrast, it is not appropriate to simply consider a zip code z and its neighboring regions $N(z)$ in isolation via a greedy approach, because for a particular zip code z, changes in capacity or pricing at some zip code that is two hops away may change our decision when optimizing over $r \ni z$. We present a simple counterexample in Figure 3.

In the counterexample provided in Figure 3, we have three zip codes where the same procedure can be obtained. Zip code A is within δ distance of B and C, but the distance between B and C is greater than δ. Nonetheless, the price of the procedure at zip code C determines how much of the capacity at B will be used in the optimal solution. For instance, at the current pricing levels, the provider in zip code B will serve all patients in zip codes A and B, while the provider in zip code B serves all patients in zip code C. However, if the price for a procedure at zip code C is decreased to 4, then the provider in zip code B serves only patients from that zip code, while the provider at zip code C serves all patients from zip codes A and C. Thus, we see that even though zip code C is not within δ distance of zip code A, the data related to zip code C still affects the optimal solution in this subset of the graph.

We can think of our optimization approach in terms of a graph problem. Let J denote the set of providers, and K denote the set of all procedures. Each provider is a node with edges representing providers in nearby regions. Each node, indexed by $j \in J$, has some level of demand for procedure $k \in K$, denoted D_j^k. Correspondingly, each node has some maximum capacity C_j^k. A pair of nodes j, j' are connected by an edge if the zip codes of their service locations are within δ of each other. Furthermore, suppose each provider j has a procedure price of μ_j^k. Our problem becomes related to a minimum vertex cover problem where we must select a collection of nodes (providers) that have the capacity to cover the demand of all neighboring unselected nodes at minimum price.

For this particular application, we focus our attention on the largest region created using $\delta =$ 50km. Since the workforce for this particular employer is largely centralized in a metropolitan area in the southern United States, it is not surprising that approximately 77% of the costs for medical procedures in this dataset fall within this largest region. Of the almost 520,000 rows related to medical procedures in the cleaned dataset, more than 400,000 occurred within this largest region. The total expenditure over the 3.5 years observed in the dataset was approximately $63.6M. In all, there were more than 21,000 providers belonging to this region in our dataset. Additionally, more than 3,600 different types of procedures were billed in this region.

3.2. Notation

Before we present our optimization approach, we introduce the notation that we use for our data, parameters, and variables.

Data:

- $k \in K$, the set of procedures;
- $j \in J$, the set of providers, each belonging to a single zip code z;
- $z \in r$, the set of zip codes in the region of interest r.
- $N(z) = \{z' \mid \|loc(z) - loc(z')\| \leq \delta, z' \in r\}$, the set of zip codes (including z itself) that are neighbors of z;
- D_j^k, the existing demand for procedure k at provider j;
- μ_j^k, the average cost for procedure k at provider j;
- C_j^k, the capacity for procedure k at provider j.

Since we are re-assigning patients in hindsight, we obtain demand information D_j^k based on the actual number of times provider j delivered procedure k in our claims dataset. We obtain the average price charged by provider j for procedure k in the same way. To protect ourselves against data error, we set a lower limit for the average cost μ_j^k based on Medicare reference pricing; if the average price for procedure k at provider j is lower than the Medicare price, we conservatively estimate that we can obtain the Medicare reference price for that geographical region.

We do not have capacity data for individual providers, so we assume that each provider can deliver up to some multiplicative factor of the historical demand from our dataset. For example, if a hospital has delivered 10 knee replacements to our patient population in the last three years, we assume that they could have delivered up to 30 knee replacements. Thus, we create a new parameter $\overline{D}^k$ that represents the maximum allowable increase under demand reallocation for a particular procedure k. We represent the capacity C_j^k as a the product $\overline{D}^k D_j^k$ for varying levels of $\overline{D}^k$. We test the sensitivity of the optimal solution to varying levels of $\overline{D}^k$ in Section 4.

We also include additional constraints that limit the total number of different types of procedures that we are willing to reallocate. This ensures that the optimal solution does not go too far in terms of reallocating all of the possible procedures that could be provided. Instead, it allows us to focus on the procedures where the maximum cost savings can be achieved. The optimal solution provides this list of procedures and could form the basis for further negotiation of prices with medical providers. We test the sensitivity of the optimal solution to varying levels of this parameter, which we denote by $\overline{\phi}$.

Parameters:

- λ, the objective penalty for each additional provider that we contract with;
- $\overline{\phi}$, the maximum number of types of procedures that are reallocated;
- $\overline{D}^k$, the maximum allowable increase under demand reallocation for a particular procedure k;
- δ, the acceptable distance between existing and new providers.

Variables:

- n_j^k, the number of patients assigned to receive procedure k at provider j;
- $\theta_{zz'}^k$, the amount of demand for procedure k transferred from zip code z to z';
- $\phi^k \in \{0,1\}$, which takes value 1 if we allow demand for procedure k to be reallocated, 0 otherwise;
- $x_j^k \in \{0,1\}$, takes value 1 if we select provider j to deliver procedure k, 0 otherwise;
- $y_j \in \{0,1\}$, takes value 1 if we choose provider j for at least one procedure, 0 otherwise.

3.3. Mixed Integer Linear Optimization Approach

Using the above notation, we formulate the following mixed integer linear optimization problem (**Formulation 1**):

$$\min_{\mathbf{n},\mathbf{x},\mathbf{y},\theta} \quad \sum_{j,k} n_j^k \mu_j^k + \lambda \sum_j y_j \qquad \text{(minimize cost + \# of providers)} \tag{1}$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{z' \in N(z)} \theta_{zz'}^k \geq \sum_{j \in z} D_j^k, && \forall k, z, && \text{(demand outflow)} && (2)\\
& \sum_{z' \in N(z)} \theta_{z'z}^k \leq \sum_{j \in z} n_j^k, && \forall k, z, && \text{(demand inflow)} && (3)\\
& \sum_k \phi^k \leq \overline{\phi}, && && \text{(limit procedure transfer)} && (4)\\
& n_j^k \leq D_j^k + M_{j,k}^1 \phi^k, && \forall j, k, && \text{(enforce procedure limit)} && (5)\\
& n_j^k \geq D_j^k - M_{j,k}^2 \phi^k, && \forall j, k, && \text{(enforce procedure limit)} && (6)\\
& n_j^k \leq \overline{D}^k D_j^k x_j^k, && \forall j, k && \text{(respect capacity)} && (7)\\
& y_j \geq x_j^k, && \forall j, k, && \text{(number of providers)} && (8)\\
& x_j^k, y_j \in \{0,1\}, && \forall j, k, && &&\\
& \theta_{z,z'}^k \geq 0, && \forall k, z, z'. && &&
\end{aligned}$$

where $M_{j,k}^1 = D_j^k(\overline{D}^k - 1)$ and $M_{j,k}^2 = D_j^k$ so that constraints (5, 6) are enforced.

The objective function (1) is to minimize the total costs incurred if we were to optimally re-assign patients to providers, with a penalty of λ for each provider that we select. This objective penalty is added so that the optimal solution does not select too many providers. From a convenience perspective, a small number of providers is desirable because it means that patients need not visit too many providers for their health care needs. Furthermore, it allows us to identify providers that have reasonable pricing for a wide range of procedures. It is notable that the solution will not necessarily select providers who have the lowest price for a particular procedure; it may select providers who have above average costs on some procedures, but low costs on others.

The constraints (2) and (3) can be thought of as flow constraints that match the original demand to the re-assigned providers via conservation of flow. Notice that the variable $\theta^k_{z,z'}$ keeps track of the amount of demand that is transferred from zip code z to neighboring zip code z'. In constraint (2), this aggregates the outflow of demand from a particular zip code z to its neighboring zip codes. In constraint (3), we consider the inflow of demand into a particular zip code z from its neighboring zip codes. Since ultimately the variable n^k_j represents the amount of procedure k that is delivered by providers j that are located within z, this constraint means that the demand flowing in must be covered.

Constraint (7) enforces capacity constraints for provider-procedure pairs, while constraint (8) allows the variable y_j to count the total number of providers that we select.

This formulation has the desirable property that it does not create too many constraints because it aggregates demand flow into the variable θ. In particular, constraints (2) and (3) need only be created for (k, z) pairs that exist in the data; that is, (provider, zip code) pairs. Since the (provider, zip code) matrix is very sparse, this greatly reduces the number of constraints that we need. We also note that even though θ is indexed by k, z, z', each individual zip code z, the size of the neighboring zip codes $N(z)$ is very small, so we need not create too many variables. Overall, even though the largest region contains more than 1,300 zip codes, more than 3,600 procedures, and over 21,000 providers, the formulation requires only approximately 177,000 constraints and 184,000 variables.

4. Insights from Computations

We implemented Formulation 1 with data from the largest region in our dataset in Julia and the mathematical programming package JuMP (Lubin and Dunning 2015). The model was solved using

Gurobi 6.5.0 as the mixed integer linear optimization solver. We ran all computations on a desktop equipped with an Intel Core i7-3770 quad-core CPU running at 3.4GHz and 32GB of RAM.

4.1. Computations Conducted

We solved Formulation 1 for the following ranges of parameter values:

- $\lambda \in \{100, 200, 500, 1000, 2000, 5000, 10000, 20000\}$, the penalty for each additional provider selected;
- $\overline{\phi} \in \{25, 50, 100, 200, 500\}$, the maximal number of types of procedures that we are allowed to reallocate;
- $\overline{D}^k \in \{2, 2.5, 3\}$, the maximal allowable increase in demand for procedure k. Where existing demand for procedure k at provider j is D_j^k, we allow the new maximal capacity to be equal to $C_j^k = \overline{D}^k D_j^k$;
- $\delta \in \{30, 40, 50\}$ km, the maximum allowable travel distance for the patient.

We set a time limit of 600 seconds for each run. Over each of the different parameter settings, we noticed that high quality solutions with optimality gap of less than 1% can be found in less than two minutes. In 240 of the parameter settings, provable optimality was achieved within the time limit. Of the remaining 120 cases, the largest optimality gap at the time limit of 500 seconds was 0.23%. Notably, all of the cases that did not reach provable optimality had large values (≥ 2000) of λ. This is due to the fact that λ implicitly defines a maximum number of providers that can be selected. Larger values of λ lead to fewer providers, making the problem more difficult to solve because the optimal solution must trade off between selecting an additional provider with minimizing overall cost. In the limit where $\lambda = 0$, the optimal solution would simply select the providers that provide each individual procedure at the lowest cost in a greedy fashion.

4.2. Computational Results and Sensitivity Analysis

We present a visualization of the overall cost reduction under all of the different parameter values in Figure 4.

In Figure 4, we have converted the different values of λ to the total number of providers that are selected in the optimal solution, which is presented on the horizontal axis. The vertical axis denotes

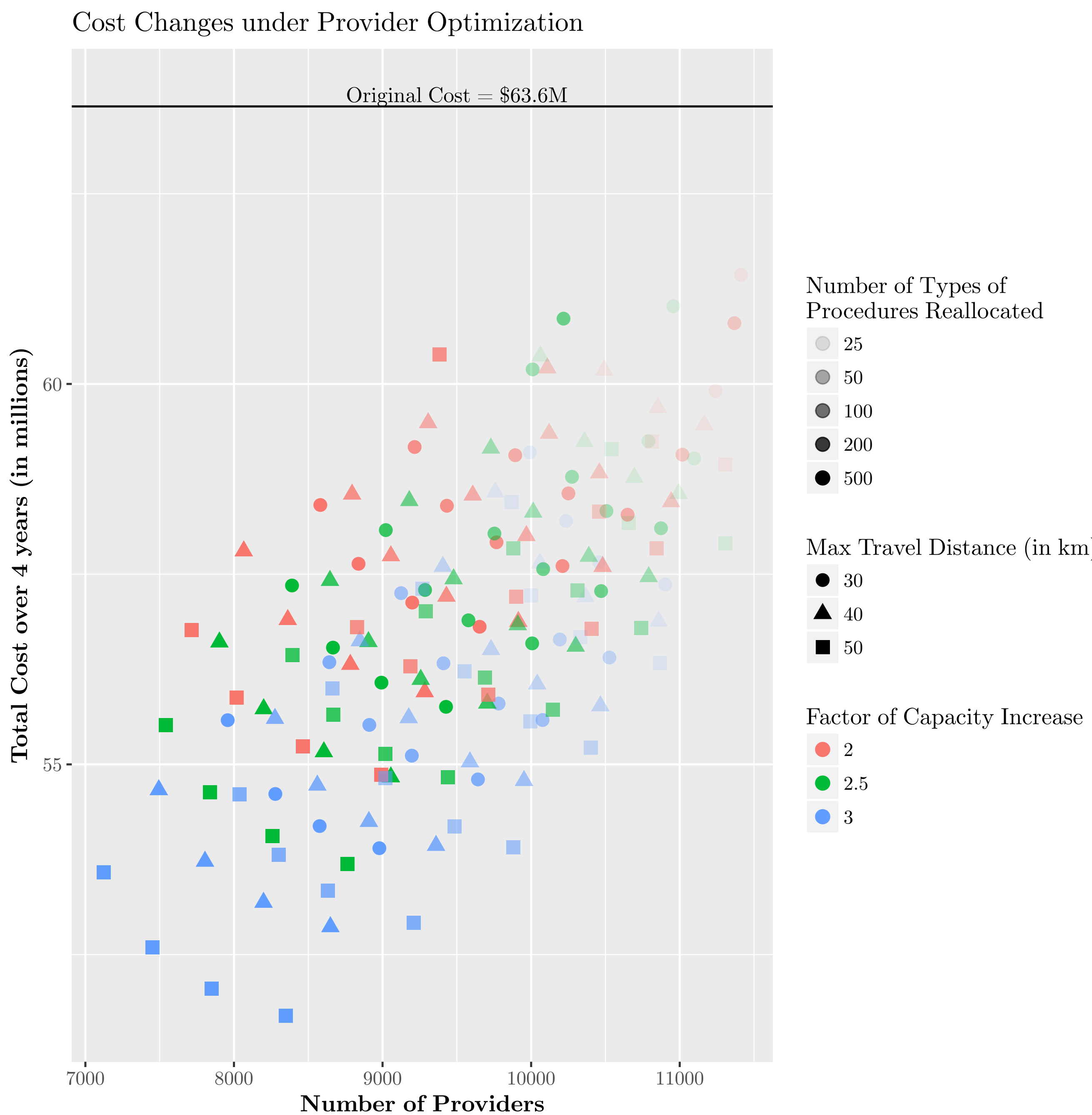


Figure 4 **Cost changes under provider optimization. Different maximum travel distances δ are denoted by different shapes; different values of capacity increase $\overline{D}^k$ are denoted by the different colors; different values of $\overline{\phi}$ are denoted by the different levels of transparency. All else being equal, larger values of δ, $\overline{D}^k$, or $\overline{\phi}$ lead to both lower cost and fewer providers.**

the total cost over the three and a half years in the database under the optimal solution provided by Formulation 1. The black horizontal line denotes the original cost of approximately 63.6 million dollars that was incurred in the data. For additional reference, more than 21,000 providers were paid for delivering procedures in the original dataset.

Each point presented in Figure 4 denotes a particular set of values for the capacity increase $\overline{D}^k$, the number of types of procedures reallocated $\overline{\phi}$, a maximum travel distance δ, and penalty λ for each additional provider. The color of the point denotes the value of $\overline{D}^k$, the level of transparency denotes the value of $\overline{\phi}$, and the shape denotes the value of δ. We see that, all else being equal, larger values of δ, $\overline{D}^k$, or $\overline{\phi}$ lead to fewer providers selected as well as lower total cost. This is an intuitive result, because larger values of δ allow for increased flexibility in selecting replacement providers. As $\overline{D}^k$ increases, we require fewer providers to deliver the same number of procedures, while also yielding additional savings from lower cost providers. Larger values of $\overline{\phi}$, or allowing more types of procedures to be reallocated, also leads to fewer providers being selected, since the optimal solution will consolidate the delivery of procedures to a smaller set of providers. It also takes additional advantage of lower cost providers for a greater number of procedures.

From this visualization, we see that under some reasonable settings of parameter values, for instance, $\delta = 40$, $\overline{D}^k = 2.5$, and $\overline{\phi} = 100$, we can achieve a total cost of approximately 57 million dollars, which would represent a savings of more than 10%. At the same time, the number of providers selected to deliver procedures is reduced by more than half. To further understand the effects of varying each of these parameters, we restrict our attention to the cases where three of the four parameters are fixed, and observe how varying the fourth parameter affects the total cost.

4.2.1. Sensitivity to Maximum Allowable Travel Distance To better understand the effects of varying the maximum allowable travel distance δ, we fix $\lambda = 1000, \overline{D}^k = 2, \phi = 100$, while varying $\delta \in \{30, 35, 40, 45, 50\}$ km. The total cost under each of these parameter settings can be see in Figure 5.

We observe that cost decreases linearly as the maximum travel distance increases. This agrees with our intuition that allowing more flexibility in finding replacement providers will yield additional cost savings.

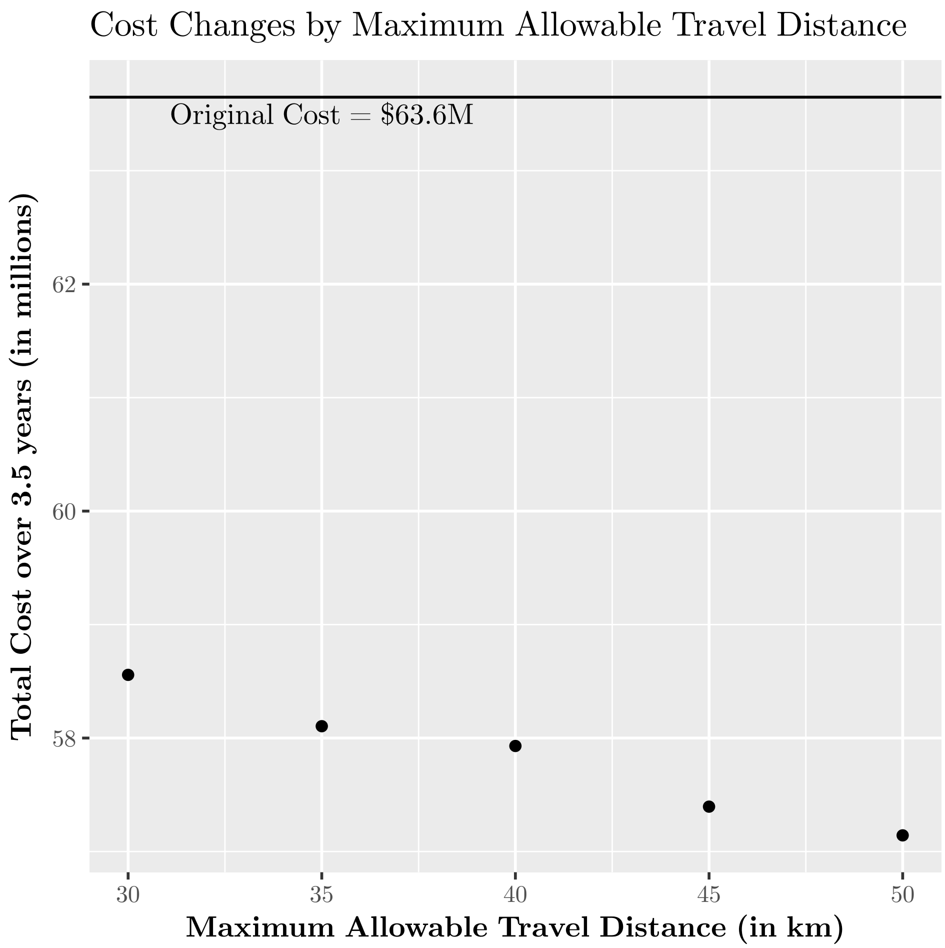


Figure 5 **Variation in total cost under different values of** $\delta \in \{30, 35, 40, 45, 50\}$ **km, with** $\overline{D}^k = 2, \lambda = 1000$**, and** $\overline{\phi} = 100$**. We observe a somewhat linear decrease in total cost as the maximum travel distance increases.**

4.2.2. Sensitivity to Number of Types of Procedures Reallocated. To better understand the effects of varying the number of different types of procedures reallocated, or the value of $\overline{\phi}$, we fix $\lambda = 1000$ and $\overline{D}^k = 2$ while varying $\overline{\phi} \in \{10, 50, 100, 200, 500, 1000\}$. The total cost under each of these parameter settings can be seen in Figure 6.

We observe that cost reductions start slowing significantly after reassigning approximately 200 different types of procedures out of a total of more than 3,600. This comports with our intuition that there are a few medical procedures on which the majority of cost savings can be realized. Our approach extracts the names of the procedures that can yield the greatest cost savings via reallocation of patients to lower cost providers. Note that the particular choice of $\overline{D}^k$ and λ do not

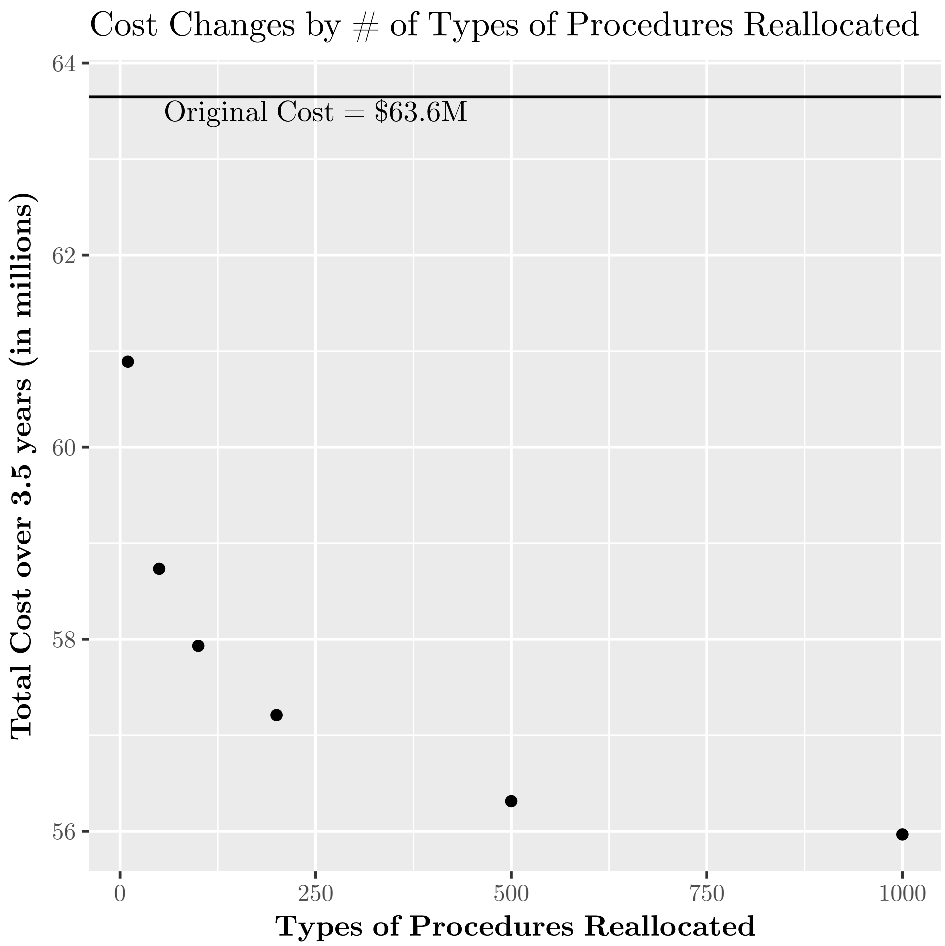


Figure 6 **Variation in total cost under different values of $\overline{\phi} \in \{10, 50, 100, 200, 500, 1000\}$, with $\delta = 40, \overline{D}^k = 2, \lambda = 1000$. We observe that cost reductions start slowing significantly after approximately 200 types of procedures being reallocated.**

affect the overall shape of this curve too much, and that fixing them allows us to better isolate the effects of only varying $\overline{\phi}$.

Increasing $\overline{\phi}$ from 500 to 1,000 yields additional cost savings of less than half a million dollars, and further increasing $\overline{\phi}$ past 1,000 leads to significantly diminished marginal savings. Since implementing these changes in provider selection are costly in practice, we would select approximately 100 to 200 different types of medical procedures on which to help patients select their providers. Even under the conservative setting of $\overline{D}^k$ and λ, we observe an approximate cost savings of 10% for values of $\overline{\phi}$ between 100 and 200.

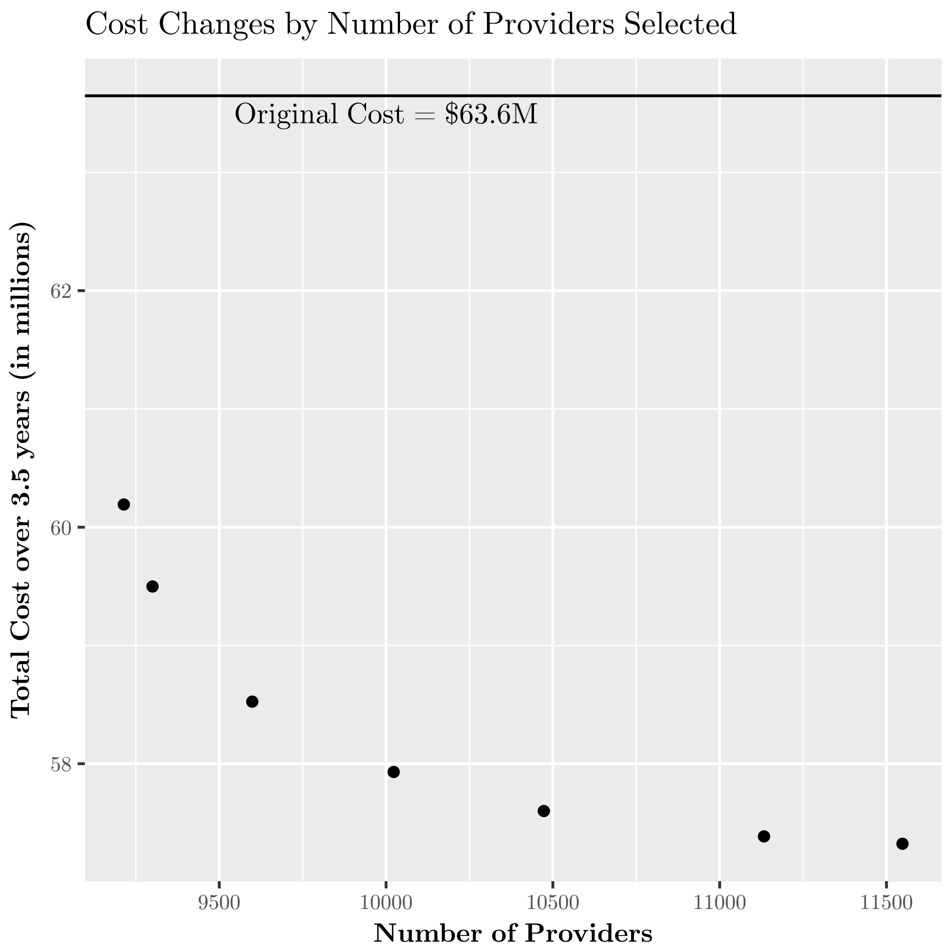


Figure 7 **Variation in total cost under different values of** $\lambda \in \{100, 200, 500, 1000, 2000, 5000, 10000\}$**, with** $\delta = 40, \overline{D}^k = 2, \overline{\phi} = 100$**. Note that each value of** λ **implicitly determines the number of providers selected. We observe that cost reductions start slowing significantly after selecting about 9,600 providers. This corresponds to setting** $\lambda = 1000$**.**

4.2.3. Sensitivity to Number of Providers Selected. To better understand the effects of varying the total number of providers selected to deliver medical procedures to our patient population, we vary the value of $\lambda \in \{100, 200, 500, 1000, 2000, 5000, 10000\}$, while fixing $\delta = 40, \overline{D}^k = 2, \overline{\phi} = 100$. Since λ is a penalty in the objective of the optimization model for each additional provider selected, each value of λ corresponds implicitly to a total number of providers. We present the cost changes under each of these different number of providers in Figure 7.

We observe that setting $\lambda = 1000$ yields a reasonable tradeoff between total number of providers selected and total cost savings. This corresponds to working with approximately 9,600 providers, which is a reduction in the original number of providers by more than half. Again, with conservative settings of $\overline{D}^k = 2$ and $\overline{\phi}$, this value of λ would bring about an almost 10% decrease in total costs related to medical expenditures in this service region.

4.2.4. Sensitivity to Capacity Assumptions.

To better understand the effects of varying the assumed additional capacity that each provider could sustain, we vary $\overline{D}^k \in \{2.0, 2.1, 2.2, \ldots, 3.0\}$ while fixing $\delta = 40, \overline{\phi} = 100, \lambda = 1000$. We plot the results in Figure 8.

Here, we see a different change in total cost for each increment in factor of capacity increase. Notice that there are two significant drops in total cost at $\overline{D}^k = 2.5$ and $\overline{D}^k = 3.0$. This is due to the discrete nature of mixed integer optimization. It suggests that for the majority of providers that we are selecting, they only provide some procedures once or twice in our dataset. We explore this further in Section 4.3.3.

4.3. Understanding Effects on Individual Procedures

We now focus our attention on the results in the case where $\delta = 40, \overline{D}^k = 2.5, \overline{\phi} = 100$, and $\lambda = 1000$. Under these parameter values, the total cost incurred is 56.8 million dollars, or approximately 11% less than the original cost of 63.6 million dollars. The total number or providers selected is slightly less than 10,000 providers, representing a 50% decrease in the number of providers used. While this aggregate information is useful for understanding the overall effects of following the prescriptions of our approach, we also investigate the effects of reallocation on prices and traveling distances for patients of particular procedures to uncover additional insight into how these cost savings are achieved. We consider the additional burden on the selected providers by considering how many more procedures they would deliver under the reallocated regime. Our analysis shows that patients need not travel far from their original service locations to yield significant cost reductions. At the same time, most existing providers are only asked to deliver a few more procedures over their current service level.

4.3.1. Changes in Pricing

The optimal solution recovers the names of providers and procedures that can yield the most cost savings overall. One particular example on which we will focus is

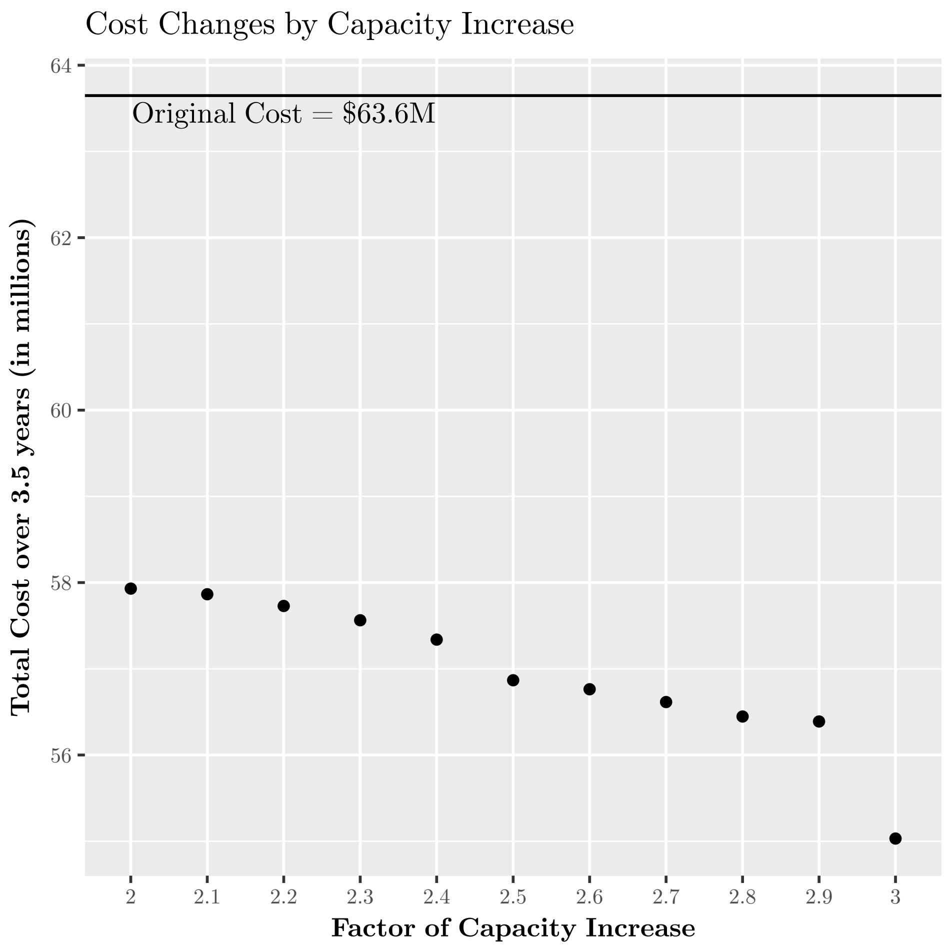


Figure 8 **Variation in total cost under different values of $\overline{D}^k \in \{2.0, 2.1, 2.2, \ldots, 3.0\}$, with $\delta = 40, \overline{\phi} = 100, \lambda = 1000$. Notice that there are significant jumps between 2.4 and 2.5, as well as between 2.9 and 3.0. This is due to the discrete nature of our optimization approach. Many of the lower-cost providers currently only deliver one or two procedures. Thus, these particular changes in $\overline{D}^k$ lead to additional cost savings. For example, allowing a capacity increase of $\overline{D}^k = 2.5$ for a provider currently delivering 2 procedures increases maximum capacity to 5. For $\overline{D}^k = 3$, providers currently delivering 1 procedure have an increased maximum capacity of 3.**

the procedure Colonoscopy/Biopsy. In this dataset, reallocating patients requiring this procedure can bring about a total of almost $129,000 in savings. In the original dataset, 284 patients received this procedure at an average cost of $1279. After optimally selecting providers, the average cost of this procedure is $825 per patient. Although we restrict our attention to Colonoscopy/Biopsy, it is important to note that similar observations can be made for many of the other procedures in

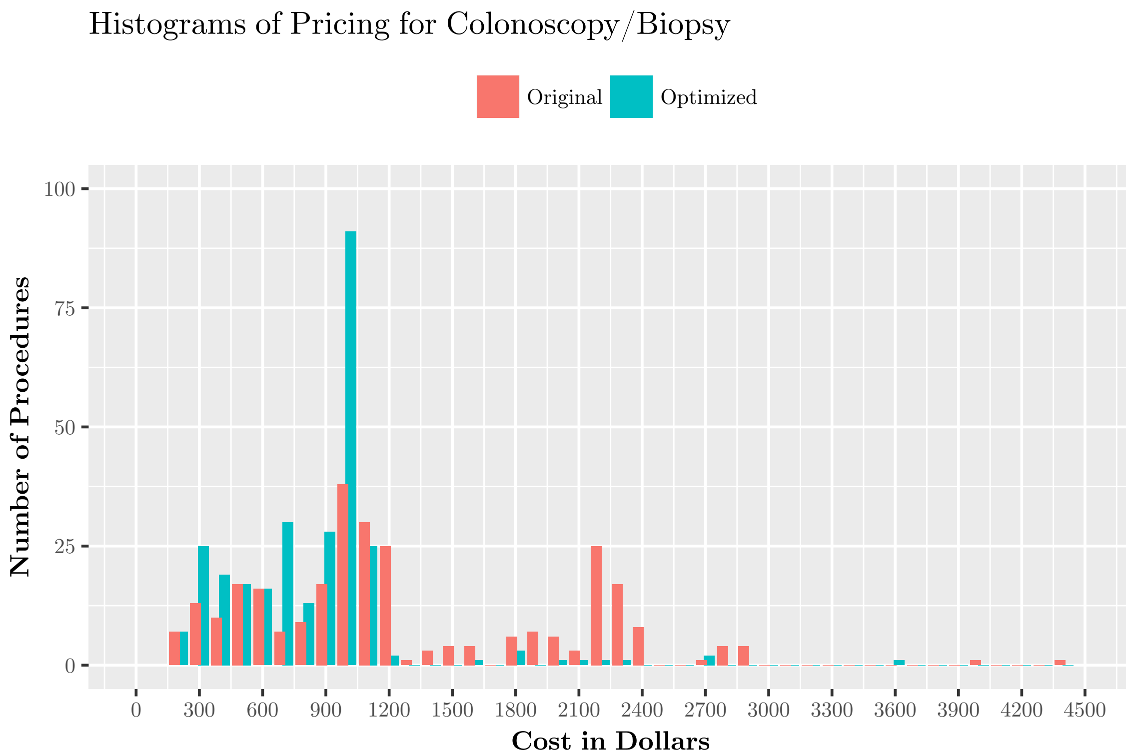


Figure 9 **Histogram of prices for colonoscopies in the original dataset and in the solution to the provider optimization problem. We observe that almost all of the procedures in our optimized regime cost less then $1,200. In the historical dataset, approximately one-third of all patients received this procedure from providers charging more than this price.**

the dataset. Figure 9 shows the distribution of the costs for Colonoscopy/Biopsy in the original dataset and in the optimal solution. We supply additional figures for other procedures in the online supplement.

From Figure 9, we can see that most Colonoscopy/Biopsy procedures are less than $1,200 in the optimized regime, while more than one-third of procedures exceeded this cost threshold in the original dataset. Under the optimal reallocation of patients to providers, we see that there is significantly less variation in prices for the procedure Colonoscopy/Biopsy. There is still some variation in price due to the geographical proximity and capacity increase constraints that we have placed on the problem. However, the optimal solution is effective at identifying costly providers and redirecting patients to providers of lower cost. It is impressive that under the reallocated regime, the 95$^{\text{th}}$ percentile of cost is comparable to the median price in the original dataset.

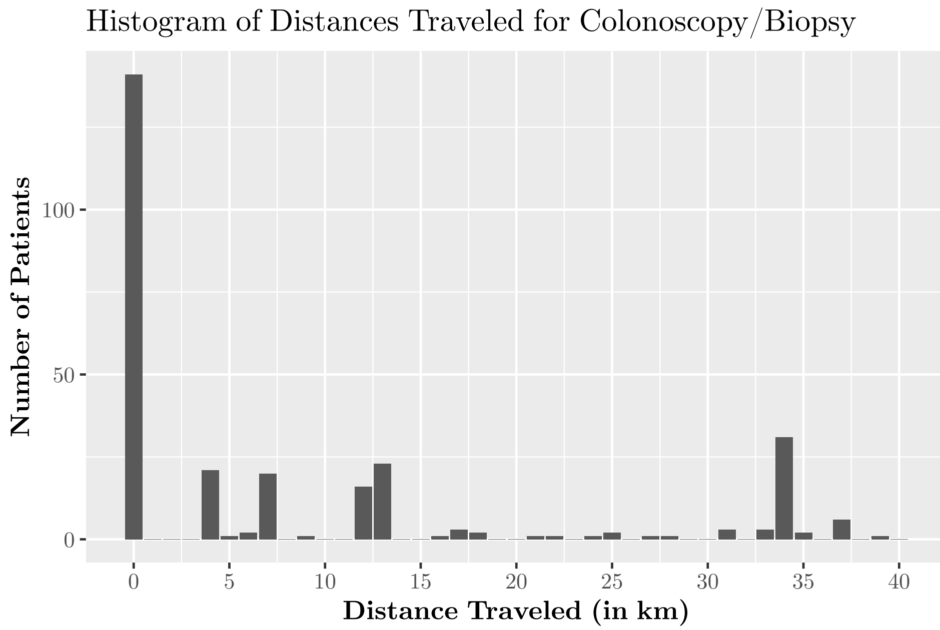


Figure 10 **Histogram of distances traveled between original service location and new provider prescribed by the optimal solution. Of the 284 patients requiring Colonoscopy/Biopsy, 141 remained in the same zip code or did not need to travel. For patients who had to travel outside of the zip code of their original service location, the median travel distance is 13 kilometers, and the mean travel distance is 18.1 kilometers.**

4.3.2. Traveling Distances The procedure Colonoscopy/Biopsy is a relatively common procedure in the dataset, so our optimization approach is able to find many lower cost providers within 40 kilometers of the original service locations. Note that not all patients must travel the maximum allowable distance under optimized regime. In fact, as we can observe from Figure 10, a majority of patients need not leave the zip code of their existing provider under the optimized regime. Of the patients who must travel outside of their original zip code, the mean and median travel distance are about 18.1 and 13 kilometers, respectively.

In practice, the parameter δ that denotes the maximum allowable travel distance can be varied. Larger values of δ would allow increased flexibility for provider reassignment, at the expense of decreased convenience to patients. To make our approach more general, we could add a penalty to the objective for each additional kilometer that a patient must travel, thus capturing this tradeoff more explicitly. In certain instances, such as very expensive procedures, it may be reasonable for patients to travel slightly farther, while for cheaper procedures, the cost savings do not warrant the additional inconvenience to the patient.

In some extreme cases, price discrepancies may be so large as to warrant treatment in an entirely different region. For instance, a group of 40 surgeons who believe strongly in price transparency opened the Surgery Center of Oklahoma, where all-inclusive prices for surgeries are posted openly on their website (Rosenberg 2013). These prices are so attractive to patients and their insurers that many patients travel across the country to receive treatment there. For the complicated procedure Lumbar Fusion, they list a price of $49,625 on their website. If the 28 patients in our dataset that received this particular procedure had sought treatment in Oklahoma, the total cost savings would be approximately $380,000. In this particular case, it may make sense for patients to incur some non-trivial cost of travel to attain more attractive pricing.

4.3.3. Changes in Capacity We now consider the impact of the reallocated patient demand from the provider's perspective. We have assumed that providers are capable of providing up to $\overline{D}^k$ times their original service capacity. Under the optimal regime with parameters $\overline{D}^k = 2.5, \overline{\phi} = 100, \lambda = 1000$, we observe that our optimal solution utilizes all of the assumed extra capacity of 2.5 times the original service capacity at 10 out of the 149 providers that delivered Colonoscopy/Biopsy in the dataset. A histogram of utilization levels by medical provider is presented in Figure 11.

Figure 11 does not tell the whole story, however. This is because many of the providers in our existing dataset only appear once. As a result, simply adding one additional patient under the optimal regime would represent a 2.0x utilization factor over current service levels. This explains the jump in cost reduction for $\overline{D}^k = 2.5$ and 3.0 that we saw in Figure 8. Figure 12 presents the absolute change in the number of Colonoscopy/Biopsy delivered across all of the providers. We see that, with the exception of 7 providers, all existing providers are asked to deliver no more than 5 additional Colonoscopy/Biopsy procedures to reallocated patients. The seven remaining providers delivered 5, 6, 7, 8, 8, 10, and 10 Colonoscopy/Biopsy procedures in the original dataset, and are asked to perform 12, 15, 15, 17, 20, 25, and 25 procedures, respectively, under the optimal solution.

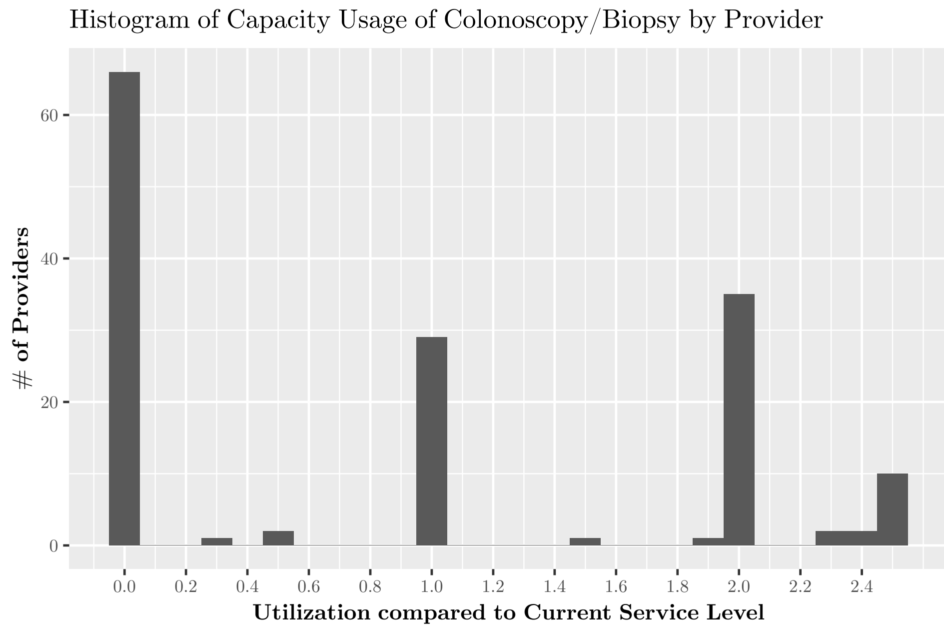


Figure 11 Histogram of capacity utilization levels at providers delivering Colonoscopy/Biopsy by multiplicative factor of original service level. Although 10 out of the 149 providers are assigned to deliver the maximum allowable number of procedures, 66 of the providers are no longer used. 29 providers remain at their current service levels.

These findings are not specific to this particular procedure. We present similar figures for other procedures such as Upper GI Endoscopy Biopsy, CT of Abdomen and Pelvis, MRI of Brain, and Lesion Removal Colonoscopy in the online supplement.

For all of the procedures reallocated in the optimal solution, we did not notice an absolute increase of more than 50 patients per year for any given procedure at any individual medical provider. This suggests to us that the assumption of a 2.5x increase in capacity utilization over the existing demand levels from our dataset is reasonable. In addition, medical providers might welcome additional demand for procedures, and may even agree to preferred pricing if minimum volume guarantees are achieved. In our view, our data-driven optimization approach could achieve cost reductions in practice by identifying specific procedures with potential for greatest savings. It also prescribes a set of low-cost medical providers within close proximity to patients' existing service locations with whom an employer could contract for even greater price reductions.

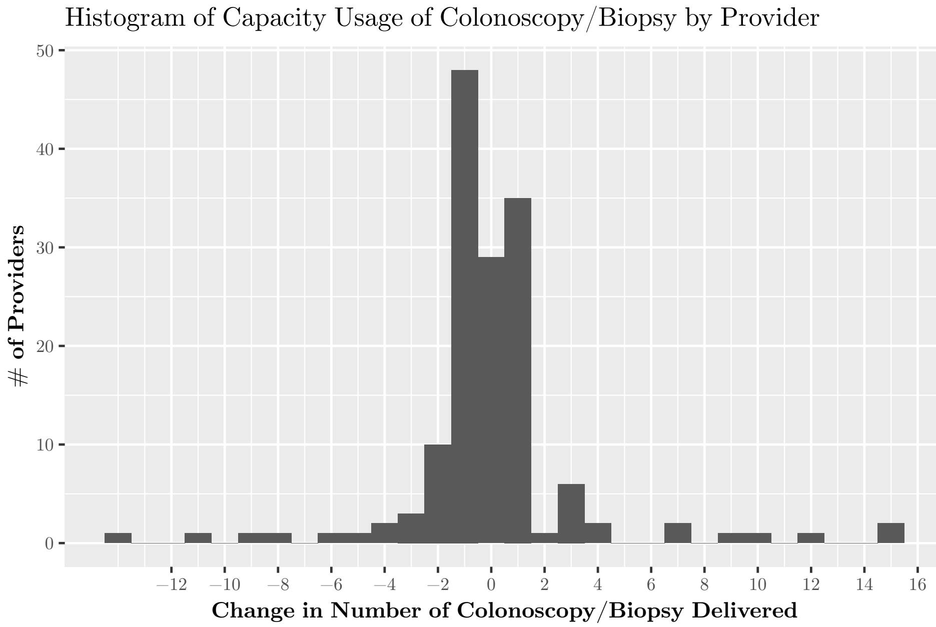


Figure 12 **Histogram of capacity utilization levels at providers delivering Colonoscopy/Biopsy by absolute change in number of procedures delivered. All except for four of the providers are asked to deliver no more than 5 additional Colonoscopy/Biopsy procedures under the new allocation. The four providers in the right-most portion of the histogram are each asked to provide the maximum allowable increase in the optimal solution.**

5. Policy Implications

Our computational results suggest significant cost reductions can be achieved through relatively simple means. Due to the large pricing discrepancies between providers delivering the same procedures within small geographical neighborhoods, patients need not be inconvenienced too much under the reassignment prescribed by the optimal solution. Indeed, our results suggest that there is ample opportunity for cost reduction; our results are just the first step in identifying all of the opportunities for decreasing an employer's expenditures in providing health care for its employees.

One immediately actionable policy would be to compile a list of low-cost providers that are capable of providing an assortment of procedures that adequately cover the desired service region. If a patient requires one of the reallocated procedures, a suitable provider in close proximity could be identified and suggested to the patient. While the patient does not necessarily need to be served by the recommended provider, incentives such as a decreased copayment or priority scheduling might convince the patient to select the recommended provider.

Another way to utilize the analysis would be to create a list of providers with whom to form service contracts. These contracts could potentially benefit both the employer and the providers: employers could benefit from decreased cost per procedure, while providers could benefit from a guaranteed minimum volume of demanded procedures.

Additionally, patients could be empowered by the findings in the data by better understanding the costs associated with their health care. While patients are somewhat immunized against the complete costs of their health care via insurance, they nonetheless are conscious of their deductibles and copayments. Providing patients with additional insight into cost variations that exist in their area would allow them to play a more proactive role in making decisions related to their health care.

6. Future Directions and Conclusion

We have presented a data-driven approach to optimally selecting health care providers based on cost and convenience to patients. This approach is easy to implement and provides approximately 10% in expenditures for medical procedures. It is important to note that this estimate is based on conservative selection of parameters in our optimization approach. In practice, relaxing some of these parameters might yield opportunities for saving up to 15% in total health care expenditures. It also provides a starting point for the selection of a smaller collection of low-cost providers with whom an employer may negotiate for even better pricing.

Many of the procedures that the optimal solution has selected are diagnostic procedures, which do not vary substantially in quality from provider to provider. This, combined with the fact that for many patients to be reassigned in the data to a particular provider we must see high existing demand, suggests that the change in overall quality for the patient population would not differ too much.

Skyrocketing health care costs in the US are unsustainable and driven in part by highly variable pricing. Using real world historical claims data, we demonstrate that a reduction in the cost of medical procedures by approximately 10% can be achieved via relatively simple means. These cost savings can be easily implemented so that medical providers are within close proximity of the patient population; furthermore, because we utilize a mixed integer linear optimization model, it is very simple to add side constraints.

We view this framework as being widely applicable to all self-insuring employers in the US as a way of mitigating the costs associated with providing health care for employees. By utilizing historical claims data to identify and select medical providers, employers can protect themselves against rampant price variation in the currently opaque market for medical procedures. Ultimately, we hope that the implementation of this framework will contribute to the sustainability of health care costs in the US.

Appendix A: Description of Data

Our dataset is 1.13GB in size, with 108 columns and approximately 1.55 million rows. Each row contains information about an insurance claim that was filed, which is described using the variables encoded in the columns. For the purposes of our analysis, we restrict attention to the following columns:

- `ClaimType`, a factor variable that denotes either medical procedure or prescription;
- `LocationType`, the type of facility where the procedure was provided (e.g., office, inpatient hospital);
- `PersonId`, a unique ID number for each individual patient;
- `DOSStart`, the start date of the medical procedure;
- `DOSEnd`, the end date of the medical procedure;
- `ICD9, Name, Category`, descriptions of the International Classification of Diseases (ICD9) code for the diagnosis;
- `CPTCode, CPTName`, descriptions of the Current Procedural Terminology (CPT) code for the procedure;
- `ProcedureName`, the name of the medical procedure;
- `ProviderName`, the name of the provider who delivered the procedure;
- `ProviderNPI`, the provider's unique National Provider Identifier (NPI);
- `AmtPaid`, the total amount paid for the procedure.

Appendix B: Data Cleaning Procedures

B.0.1. Identifying Procedure Delivery Locations. An issue specific to claims datasets is that locations within the dataset often refer to billing addresses, as opposed to the actual location where medical services were delivered. Using an additional dataset provided by the Centers for Medicare and Medicaid Services that specified the locations of medical providers based on their National Provider Identifier (NPI), we were able to obtain the actual provider location (Medicare 2015). Although this allowed us to properly identify the zip codes of practice locations, many of the rows in our dataset did not have valid NPI information. There was no additional information in the dataset from which we could obtain the practice locations of these claims, so we discarded them for the purposes of our analysis. Of the original 1.5 million rows, we discarded 100,000 rows (approximately 6%). These removed rows accounted for approximately 10% of the cost in the entire dataset.

One additional challenge is that a single procedure might generate multiple claims with different delivery locations. Upon further examination, we noticed that for almost all of these instances, one of the claims was billed by a physician or specialist, while the other was billed by a hospital or medical center, indicating that a medical provider was utilizing nearby hospital facilities. To obtain the correct delivery location, we concluded that the actual delivery zip code of the medical procedure was in the hospital or medical center. We were able to perform this location finding procedure by utilizing the `LocationType` column in the dataset and specifying a precedence ordering over the location types (e.g., outpatient hospital takes precedence over office).

It is important to note that this data cleaning step is also crucial in determining the true cost distribution for a specific provider-procedure pair, which we describe below.

B.0.2. Obtaining Costs for Provider-Procedure Pairs. Two challenges involved in obtaining an accurate distribution of costs for all provider-procedure pairs are procedures involving multiple providers and refunded claims. As we discovered while identifying procedure delivery locations, a single procedure provided to a single patient can often result in multiple claims. Thus, it is not appropriate to aggregate rows simply by `ProviderName` and `ProcedureName`, because multiple providers may have taken part in the same procedure. We instead identify all of the instances in which a single procedure led to multiple claims and we aggregate these rows individually. When we aggregate these rows, we replace them with a single row containing the sum of all of the costs paid for that procedure. We also create a new provider that is a combination of all of the providers who took part in the procedure. As before, we use location type precedence ordering to determine the true delivery location of the procedure.

References

Beck, Melinda. *How to Bring the Price of Health Care Into the Open*, The Wall Street Journal. 23 February 2014. Retrieved January 20, 2016.

Centers for Medicare and Medicaid Services. National Plan and Provider Enumeration System (NPPES) Data Dissemination. Retrieved January 20, 2016.

Commonwealth of Massachusetts. Recommendations of the Special Commission on Provider Price Reform. 9 November 2011. Retrieved January 20, 2016.

Emanuel, Ezekiel, et al. *A Systemic Approach to Containing Health Care Spending*, The New England Journal of Medicine. **367;10** (2012), 949–954.

Lubin, Miles, and Iain Dunning. "Computing in Operations Research using Julia". INFORMS Journal on Computing, **27;2** (2015), 238-248.

Reinhardt, Uwe E. *The Disruptive Innovation of Price Transparency in Health Care*, Journal of the American Medical Association. **310;18** (2013).

Rosenberg, Tina. "Revealing a Health Care Secret: The Price". The New York Times. Retrieved February 21, 2017.

Sinaiko, Anna D., and Meredith B. Rosenthal. *Increased Price Transparency in Health Care – Challenges and Potential Effects*, The New England Journal of Medicine. **364;10** (2011), 891–894.

Squires, David A. *Explaining High Health Care Spending in the United States: An International Comparison of Supply, Utilization, Prices, and Quality*, The CommonWealth Fund. **10** (2012).

Stanton, Mark W. *The High Concentration of U.S. Health Care Expenditures*, Agency for Healthcare Research and Quality: Research in Action. **19** (2005).