IISE TRANSACTIONS ON HEALTHCARE SYSTEMS ENGINEERING
https://doi.org/10.1080/24725579.2019.1567629

A multi-objective model for optimizing staffing across geographically distributed patient-centered medical homes

David Linz, Zelda B. Zabinsky, Joseph Heim, and Paul Fishman

Industrial Engineering, University of Washington, Seattle, WA, USA

ABSTRACT

As demand for medical services increases, it may become necessary to centralize provision of medical care, particularly specialty services, to certain locations within a geographic region. Due to tradeoffs between operating costs, the risk of limited access, and decreased availability of patient-centered care, there is a need to help healthcare administrators make informed decisions. This article examines the issue of staffing clinician care in a region of locally distributed patient-centered clinics. As previous literature suggests, there are some benefits, as well as drawbacks, to centralizing the provision of care. Therefore, an administrator must understand the tradeoffs between the risk of not meeting patient demand while considering staffing expenses, patient travel time, and lack of continuity with large centralized clinics. We propose a multi-objective mixed-integer program to minimize the risk of insufficient staffing as well as minimize an aggregated penalty function that incorporates staffing expenses, patient travel time, and a discontinuity penalty term. The methodology provides an efficient frontier of risk versus penalty and we demonstrate the approach with numerical results using data sampled from a typical demand distribution. Furthermore, the numerical examples demonstrate how optimal decisions could vary, depending on the distribution that characterizes demand, particularly when the demand distribution has non-normal or heavy-tailed effects.

ARTICLE HISTORY
Received 17 December 2015
Accepted 18 November 2018

KEYWORDS
Healthcare management; integer programming; outpatient clinics; risk management; stochastic programming

1. Introduction

Modern healthcare management struggles with staffing medical services in a way that satisfies patient demand while providing patient-centered care that makes an efficient use of resources in a healthcare network (Smith-Daniels *et al.*, 1988; Rais and Viana, 2011). The specific staffing choices that health systems have (e.g. the ratio of primary to specialty care providers and the division of staff among clinic locations) are driven by a number of factors that include: the health care needs of the population, the relative supply of physician and non-physician personnel in the community, and the financial cost of hiring additional staff. Although specialist location is influenced by reimbursement models and individual provider strategies, there is no consensus within research communities about the optimal strategy to staff clinics or how to decide on a provider mix that achieves the best outcomes within a patient-centered framework (Smith-Daniels *et al.*, 1988; Rais and Viana, 2011). Moreover, variation in patient demand often leaves clinicians unable to treat patients promptly (Bodenheimer and Pham, 2010).

Clinics face the challenge of paying for excess capacity during low demand periods and, worse, having unmet or inadequately addressed patient needs during high demand periods (Smith-Daniels *et al.*, 1988). This challenge is exacerbated when patient care requires interaction among multiple specialists. The problem of highly variable demand is compounded when the demand for specialty service exhibits high kurtosis or "heavy" tails (e.g. when rare or infrequent events drive a significant amount of the overall weekly specialty demand). We consider weekly specialty demand distributions with heavy tails and quantify the risk of insufficient staffing during demand spikes.

One approach to handle high variation within a population's need for specialty care is to redirect patients to large facilities that are able to handle fluctuations in demand. In these circumstances, patients typically served by local neighborhood clinics are redirected to large healthcare centers for certain types of specialty care. Similarly, weekly staffing levels associated with those specialty care types are allocated to those large healthcare centers to serve the combined neighborhood demand.

Under this centralized provision of care, large specialty centers with expanded panels of patients would have less risk of being under- or over-staffed. Moreover, a centralized provision of care benefits from economies of scale, including sharing supplies and administration expenses while reducing overhead and increasing care coordination (Luft and Crane, 1980; Cohen and Lee, 1985).

CONTACT David Linz ddlinz@uw.edu Industrial Engineering, University of Washington, 3900 Northeast Stevens Way Mechanical Engineering Building, Room G6, Seattle, WA 98105, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhse.

Centralizing provision of care may also have significant drawbacks. Increased transportation times could be prohibitively high for some patient communities and could present a significant burden to provision of care. Also, centralized provision of care could have large impacts on the "continuity" patients experience, as some patients will be much less likely to be scheduled with the same specialist in large clinics than in smaller clinics closer to their own neighborhoods (Saha *et al.*, 2008). Patients may also be unfamiliar with a large clinic, which could discourage them from seeking the specialist care they need.

An administrator of a healthcare network must, therefore, weigh the advantages and disadvantages of centralized versus distributed staffing. Their considerations would include: (1) patient's reliable access to specialists in a timely fashion; (2) direct cost of staffing; (3) increased burden on patients traveling for care; and (4) lack of familiarity and possible discontinuity of care that patients experience in large centralized clinics.

Objectives fall into two types: those that pertain to average costs incurred by the clinics and patients, and the objective to avoid under-staffing when demand spikes occur. For instance, both patient travel time and staffing costs are experienced every day and, therefore, can be measured well by weekly averages or expected values. Conversely, ensuring reliable access to care in a timely manner requires a policy that avoids circumstances where a spike in patient demand for a week exceeds the capacity of staff at a clinic. This objective is better represented with a "risk" metric that emphasizes the infrequent events in the tail of the demand distribution.

To address these two different types of considerations, we develop a model that considers two objective functions and borrows methods popular in portfolio analysis (where average return is optimized relative to risk of loss). Our model uses a weighted combination of weekly average "penalty cost" along with a "risk" measurement that captures consequences from the entire weekly demand distribution. This formulation with two objectives should give an administrator an effective tool to weigh the risk of being understaffed with the average cost of operation. To aid in such a decision, this article proposes a multi-objective mathematical optimization model that balances the tradeoffs between a risk function and a penalty function.

There are several possible risk measures that measure aversion to being understaffed relative to patient demand. These common risk measures include: expected value, quantile or "Value-at-Risk" (VaR), and the "Conditional Value-at-Risk" (CVaR). The expected value provides a measure of central tendency, which in the context of staffing represents the average number of weekly hours of demand that exceed staffed hours. Our model includes a constraint to ensure average weekly demand is met. However, exclusively using averages in this situation may not provide enough protection against being understaffed.

The quantile metric VaR provides more information about extreme behavior at some given level γ. For example, considering a distribution of *unmet* demand, X, if the VaR measure at level $\gamma = 0.20$, equals zero, $VaR_{0.20}(X) = 0$, then demand is satisfied 80% of the time, and not satisfied 20% of the time. However, a quantile metric does not capture the impact of scenarios beyond the quantile level. In these circumstances, a given quantile will ignore significant impact of scenarios beyond the quantile level.

The CVaR metric addresses the issues of both the expected value and the quantile metric, since it measures the *amount* of unsatisfied demand in scenarios beyond a quantile level. For example, given the previous distribution of *unmet* demand, X, if $CVaR_{0.20}(X) = 100$, then in the 20% of time demand is not met, there is an average of 100 h of unsatisfied demand. This measure allows decision makers to consider scenarios where high demand exceeds staffed time by a significant number of hours. The CVaR measure of risk is conservative in nature, which may be appropriate for risk-averse decision makers. The measure also has useful properties (e.g. coherence and convexity) for optimization (Rockafellar and Uryasev, 1999). The risk measure CVaR has been used in sample average approximation for stochastic optimization (Kleywegt *et al.*, 2002; Wang and Ahmed, 2008). CVaR is a popular way to quantify the risk of poor outcomes and has been used widely in financial applications (Alexander and Baptista, 2004; Zhu and Fukushima, 2009; Yu, 2011) as well as medical decision making (Zheng *et al.*, 2015).

We use CVaR to measure the amount of demand per week that exceeds the specialist capacity staffed. Given that decision makers will consider this risk measurement along with a general concern for average system performance, the problem is posed with two objective functions: the first objective is a sum of CVaR over all clinics characterizing the number of unsatisfied weekly demand hours; the second is a "penalty" function that combines the scaled linear cost of staffing, patient travel time, and patient discontinuity factor for large centralized facilities. Using the proposed model, a decision maker can assess the tradeoffs implicit in locating specialty care and examine the impact of heavy-tailed demand distributions on the optimal staffing recommendations.

We propose and solve a multi-objective stochastic optimization problem that incorporates both the location of the specialist staff and the staffing levels (in hours per week) to account for the combined risk interactions. We contrast our approach with a deterministic model, which is a common method to determine the location and demand fulfillment decisions. We compare the risk exposure (in terms of unmet demand) of a solution generated by the deterministic formulation with our multi-objective stochastic method and demonstrate that our handling of specialist location and demand fulfillment leads to lower risk exposure than the deterministic method. We also present solutions on the efficient frontier to illustrate tradeoffs between the penalty formulation and risk with the multi-objective stochastic method.

2. Background

Staffing healthcare systems is a critical issue in the management and provision of cost-effective medical services

(Gaynor and Anderson, 1995; Li and Benton, 2003; Abri *et al.*, 2006; Jack and Powers, 2009). Reviews (Smith-Daniels *et al.*, 1988; Rais and Viana, 2011) describe various optimization models directed at improving hospital capacity management and emphasize the importance of strategies to reduce costs. In particular, Smith-Daniels *et al.* (1988) point to the importance of economies of scale within networks of healthcare facilities and the benefits of consolidating services within geographic regions to eliminate redundancy and optimally locate staff and resources. Following Smith-Daniels *et al.* (1988), our article provides a data-driven numerical model that addresses the question of optimal staffing with respect to care consolidation.

The benefit of consolidating care to improve efficiency has been well established in the literature (Parker and DeLay, 2008; Jack and Powers, 2009; Sikka *et al.*, 2009; Trinh *et al.*, 2014). Benefits of consolidation include removing redundancy, economies of scale, and increased communication between specialists. Other research demonstrates that consolidating patient volume can result in improved healthcare quality and health outcomes (Halm *et al.*, 2002; Huckman and Pisano, 2006). Rohleder *et al.* (2006) present a model that demonstrates the advantage of care-consolidation to reduce the impact of uncertainty in patient demand and the risk of disruption. Reducing uncertainty by pooling patients together into large groups, also referred to as "risk-pooling," is a key benefit to consolidating care.

Another critical element to providing efficient, patient-centered medical care is locating professionals and resources close to patients. Research has shown the importance of locating care provision near patients (Kohli *et al.*, 1995; Payne *et al.*, 2000; Jordan *et al.*, 2004; Woods *et al.*, 2005; Exworthy and Peckham, 2006; Cook *et al.*, 2007; Saha *et al.*, 2008). There is a large literature on optimal medical facility location (Calvo and Marks, 1973; Shuman *et al.*, 1973; Oliveira and Bevan, 2006; Griffin *et al.*, 2007; Baray and Cliquet, 2013) (for a review, see Daskin and Dean (2004)), which broadly applies facility location models to a medical context.

The use of risk measures to optimize the provision of service and care has been shown to be effective. Value-at-Risk (VaR) is used as a risk measure, incorporated as chance constraints, when optimizing quality of service across networks (Shen and Chen, 2013). Furthermore, the incorporation of a risk-based perspective (through chance constraints again), is used in stochastic programing models for setting nurse-to-patient ratios (Maass *et al.*, 2017). Chen *et al.* (2006) used a measurement of risk based on the "Mean Excess Regret" or CVaR for facility location, which is the same risk measurement we examine for allocating staffing resources.

A multi-stage stochastic program with recourse (Birge and Louveaux, 1997; Shapiro *et al.*, 2009) is an optimization approach that accounts for uncertainty. The models in Maass *et al.* (2017) and Shen and Chen (2013) are formulated as two-stage stochastic programs with recourse, where the first stage decision is at strategic level, and the second stage is at an operational level once the uncertainty has been revealed. Our model considers both location and staffing levels of specialists at a strategic level, so we do not have a recourse decision at the operational level. It is also common for a multi-stage stochastic program with recourse to merge cost considerations into a measure of risk. In this article, we keep the penalty measure (including patient travel time, lack of familiarity and possible discontinuity of care, and direct cost of staffing) separate from the risk measure (the CVaR of understaffing and contributing to lack of access to care). By considering both objectives, we can explore solutions along an efficient frontier to aid a decision maker in understanding tradeoffs between penalty and risk. We also demonstrate the impact that probability distributions of demand with heavy tails versus light tails have on the decisions and subsequent risk of under-staffing.

Less research exists concerning the allocation of resources inside of existing clinic facilities. Deterministic models, such as Ruth (1981), explore the possibility of allocating beds within a group of facilities while Stummer *et al.* (2004) address the location of capacity within pre-existing facilities, using a multi-objective optimization problem to determine a location and size of medical departments relative to service objectives. Benneyan *et al.* (2012) use a deterministic mathematical model to optimize staffing inside a geographically distributed network of facilities. In Benneyan *et al.* (2012), a decision maker makes a series of staffing and demand allocation decisions to minimize a combination of non-coverage cost, staff cost, and travel cost.

An approach to locating specialist staff that addresses stochastic demand is proposed by Mahar *et al.* (2011). Assuming normally distributed demand, Mahar *et al.* (2011) develop an integer program based on a weight penalty for expected unmet demand and a cost for transportation and staffing at each potential location. Mahar's model improved upon more basic location based models (such as set-covering) by accounting for economies of scale and a stochastic model of demand. Using separate terms for flexible and inflexible demand (demand that cannot be moved to another facility), Mahar *et al.* (2011) were the first to rigorously demonstrate how specialty care can be optimally allocated to benefit from resource pooling with minimal transport costs. The model described by Mahar *et al.* (2011) is limited to normally distributed demand; moreover, the model is restricted to minimizing *expected* costs and does not directly deal with patient demand that will naturally follow the re-location of other specialist services. In this article, we expand the perspective to include a measure of risk that emphasizes tail-effects that might not be otherwise captured when examining expected cost of staffing shortfall under normally distributed demand as well as describe demand, created by patient co-morbidity.

There are advantages to using a model for staffing that takes into account tail-effects and non-normal demand. Without these considerations, there is a distinct possibility that a solution may underestimate the needed staffing in clinics with infrequent but critical spikes in demand. In

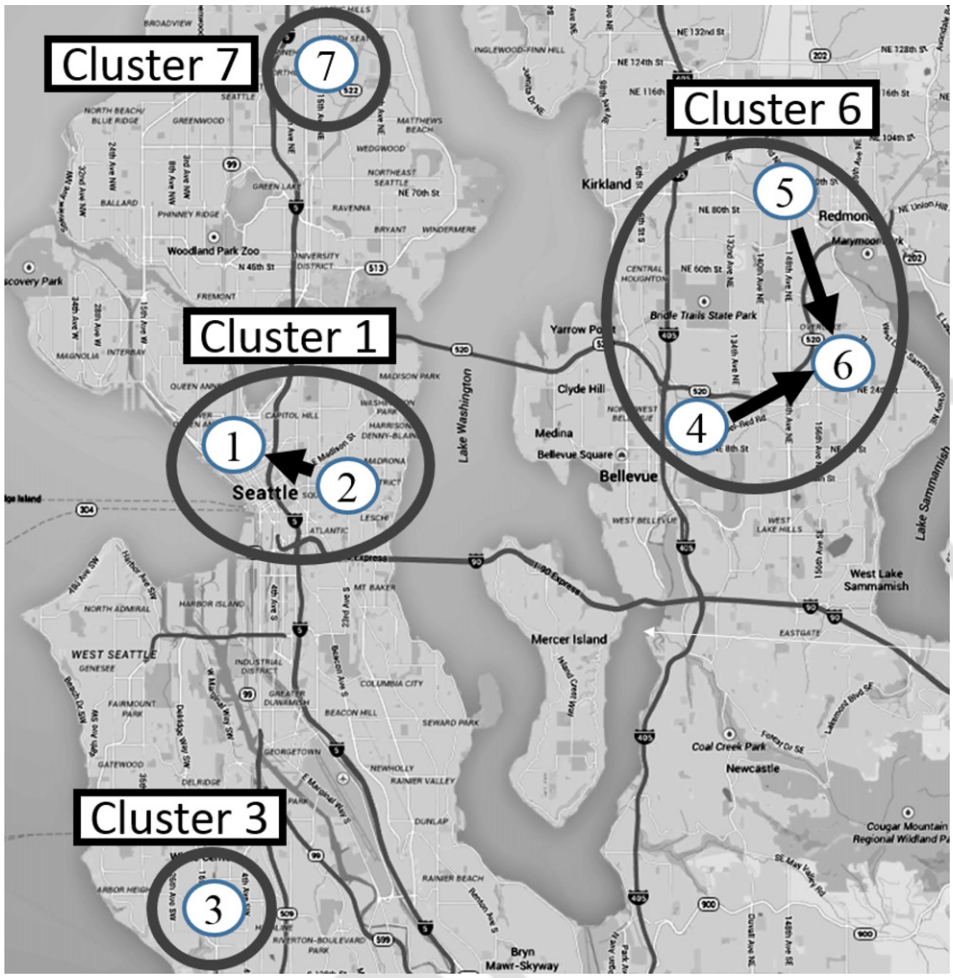

Figure 1. An example of seven clinics in the greater Seattle area, marked with circles and labeled 1–7. Clusters where specialist care is shared between clinics are marked with the shaded circles. The arrows indicate the central clinic inside the cluster that provides specialist care.

some circumstances, medical demand has been shown to have high kurtosis and effects that are dominated by extreme values (Mihaylova *et al.*, 2010). In these cases, it is important to use a model that takes into account the effects of distributions that are not normal. In absence of a model that can measure the effects of extreme events, there is a chance that an optimization model will provide staffing strategies based on an underestimation of the risk of insufficient staff. Therefore, an optimization model that does not account for non-normal or heavy-tailed distributions in demand would provide sub-optimal solutions to a risk-averse decision maker.

This article proposes a multi-objective optimization model that does not rely on the assumption of demand normality (or any specific demand distribution). Instead, our approach employs data to model the risk due to tail effects for a general distribution. With a measurement of risk based on CVaR, as in Chen *et al.* (2006), our article introduces a method for optimizing the specialist care staffing among geographically located patient-centered clinics. The results allow a decision maker to understand the risk and penalty tradeoffs, even in the case where demand distributions exhibit non-normal behavior. Decisions could, therefore, be better made with regard to spikes in demand that might nonetheless impact a staffing strategy's optimality.

3. Optimization model

Our model optimizes staffing within a geographic region with multiple patient-centered healthcare outpatient clinics where specialists might be located.

This research was motivated in part by questions raised by the leadership of Group Health Cooperative, an integrated health system that operates primarily in the Puget Sound region of western Washington, which includes metropolitan Seattle. Group Health leadership has considered various ways to organize its care system to meet patient demand for all services, from co-locating primary and specialty providers in the same facility to creating larger, centralized specialty centers. This may involve re-balancing staffing hours inside existing clinics that staff care or re-grouping clinics into different clusters that share the provision of care at one central location.

For example, seven pre-existing clinics in the greater Seattle area (labeled 1–7) are shown in Fig. 1. Patients typically use their nearest clinic for primary care and may either obtain specialty care locally, or at another clinic where the required type of specialist is staffed. We consider the administrative decision of determining the optimal way to staff specialists in various locations. The example in Fig. 1 has four clusters: Cluster 1 includes demand from both clinics 1

Table 1. System indices, parameters, decision variables, and random variables.

Indices	
C	Number of clinics, indexed by $c = 1, \ldots, C$
C	Number of potential clusters, indexed by $l = 1, \ldots, C$
I	Number of specialty types, indexed by $i = 1, \ldots, I$
Parameters	
$m_{i,i'}$	Fraction of demand of specialty type i that should be served at the same location where specialty type i' is provided
$h_{l,i}$	Penalty cost of staffing one hour of specialty type i at cluster l
$t_{c,l,i}$	Penalty cost of transporting an hour of demand of type i between clinic c and cluster l
$v_{l,i}$	Threshold for a large cluster l for specialty type i
$s_{l,i}$	Penalty for exceeding threshold for a large cluster by one hour at cluster l for type i
w_i	Scaling index for risk-aversion from specialty type i
$b_{l,i}$	Maximum capacity at each cluster l for specialty type i
γ_i	Threshold for risk for specialty type i
Decision variables	
$X_{c,l,i}$	Binary variable indicating if clinic c is assigned to cluster l and offers specialty service of type i.
$Y_{l,i}$	Real variable showing the hours of specialty type i staffed at cluster l
Random variables	
$D_{c,i}$	Random variable representing the demand at clinic c for specialty type i, with mean $\mu_{c,i}$ and random samples $d_{c,i,k}$ for $k = 1, \ldots, K$

and 2 with specialists located at clinic 1; Cluster 3 consists only of clinic 3, which provides its own specialist care; Cluster 6 handles demand from clinics 4, 5, and 6 with specialist care provided at clinic 6; Cluster 7 consists of clinic 7, which provides its own specialist care.

Our model assigns weekly specialist hours to staff the clinic that serves each cluster (our convention is that the cluster number will take the name of the supporting clinic). By solving both the clustering problem and the service assignment problem in the same optimization model, we address the considerations of healthcare management.

The model also takes into consideration the interaction between different types of specialists as some specialties must be co-located. For example, a portion of endocrinology patients will also need oncology care; therefore, there must be sufficient allocation of endocrinologists and oncologists at the same clinic to serve patients needing both types of care.

The model makes use of quantifiable patient "demand" for specialty services at each clinic. We measure this demand as the time required to serve all patients who use that clinic and are seeking care within a given time frame (e.g. weekly). Demand has a corresponding probability distribution over a given period of time. Based on this observed demand, an administrator will want to optimally cluster and staff specialists throughout a set of clinics.

The model assigns each clinic to a cluster for each specialty type. Since there may be priorities in how different specialties are centralized (e.g. patients needing both oncology *and* endocrinology will prefer to have endocrinology located with their oncologist and not vice versa), the model imposes a "hierarchy" such that clinics that act as centers for one type of specialist care will have to provide specialist care of a lower priority. Staffing levels have to take into account the additional demand attributed to "co-morbidity" (e.g. demand coming from patients being referred to that clinic for another type of care who would also like to receive care of the lower priority type).

Since the model is concerned with the strategic decision to consolidate care, the method focuses on staffing *specialist hours* rather than the individual specialists. This perspective allows decision makers to look at the strategic allocation of staffing *resources* rather than the more operational problem of scheduling a group of professionals.

3.1. Model formulation

The key task of the model is to determine the optimal staffing allocation that minimizes specified functions for risk and penalty. This involves choosing the location of clinics within clusters as well as selecting the number of staff hours per cluster.

We consider C clinics where each clinic is indexed by c, $c = 1, \ldots, C$. Let there be C potential clusters indexed by l, $l = 1, \ldots, C$. We model I healthcare professional types indexed by i, $i = 1, \ldots, I$.

The decision variables are described relative to the specified indices. To represent the assignment of clinic c to cluster l, for each specialty type i, let

$$X_{c,l,i} = \begin{cases} 1, & \text{if clinic } c \text{is incluster } l \text{ for specialty type } i \\ 0, & \text{otherwise} \end{cases}$$

for $c = 1, \ldots, C, l = 1, \ldots, C$ and $i = 1, \ldots, I$. Note that for $X_{c,l,i} = 1$, if $l = c$ then clinic l staffs specialists for itself and the other clinics assigned to cluster l.

Let $Y_{l,i}$ be a decision variable specifying the number of hours per week of specialist type i staffed at cluster l, for $l = 1, \ldots, C$ and $i = 1, \ldots, I$.

The optimization problem depends on one basic random vector $D_{c,i}$ which represents the weekly demand at clinic c for service type i. The probability distribution for $D_{c,i}$ is allowed to be a general distribution, as long as it has finite mea; e.g. $E(D_{c,i}) = \mu_{c,i} < \infty$. The probability distribution may be based on an empirical model of demand. Although a closed form cumulative distribution function is not required, the optimization model assumes that the demand distribution $D_{c,i}$ has a computable $\mu_{c,i}$ as well as a method for generating an arbitrary number of random samples denoted $d_{c,i,k}$ for $k = 1, \ldots, K$.

The decision variables, random variables, and parameters are summarized in Table 1.

The optimization problem in (1)–(9) is formulated as a two-objective mixed integer program. Section 3.2 discusses how we construct the efficient frontier for the two objectives

using the ϵ-method (Deb *et al.*, 2016). Section 3.3 details how we approximate the risk function, $Risk_{\gamma_i}$ written in terms of the random variable $D_{c,i}$, with K demand samples $d_{c,i,1}, \ldots, d_{c,i,K}$ from the demand distribution for $D_{c,i}$:

Minimize:

$$\sum_{i=1}^{I}\left(w_i \cdot Risk_{\gamma_i}\left[\sum_{l=1}^{C}\left(\sum_{c=1}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right)^{+}\right]\right) \quad \text{Risk (1)}$$

$$\sum_{i=1}^{I}\left(\sum_{l=1}^{C} Y_{l,i} \cdot h_{l,i} + \sum_{c=1}^{C}\sum_{l=1}^{C}\left(\sum_{i'}^{I} m_{i',i} \cdot \mu_{c,i'} \cdot t_{c,l,i'}\right) \cdot X_{c,l,i} + \sum_{l=1}^{C}[Y_{l,i} - v_{l,i}]^{+} \cdot s_{l,i}\right) \quad \text{Penalty (2)}$$

Subject to:

$$\sum_{c=1}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot \mu_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i} \leq 0 \quad \forall l, i \quad \text{Average demand satisfaction (3)}$$

$$Y_{l,i} \leq b_{l,i} \quad \forall l, i \quad \text{Capacity (4)}$$

$$\sum_{l=1}^{C} X_{c,l,i} = 1 \quad \forall c, i \quad \text{Complete allocation (5)}$$

$$X_{l,l,i} - X_{c,l,i} \geq 0 \quad \forall c, l, i, \ s.t. \ c \neq l \quad \text{Clusters provide own care (6)}$$

$$X_{l,l,i'} - X_{l,l,i} \geq 0 \quad \forall l, i, i' \ s.t. \ i' > i \quad \text{Hierarchy for centralizing specialist types (7)}$$

$$X_{c,l,i} \in \{0, 1\} \quad \forall c, l, i \quad (8)$$

$$Y_{l,i} \geq 0 \quad \forall l, i \quad (9)$$

The first objective function (1) characterizes a linear combination of the risk of patient demand exceeding the staffed hours at clusters for each specialist type. Here, $Risk_{\gamma_i}$ represents $CVaR_{\gamma_i}$, as detailed in Section 3.3. The argument of the risk function, also called the "loss function," is

$$\sum_{l=1}^{C}\left(\sum_{c=1}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right)^{+}$$

and makes the use of the positive portion of the argument to determine unmet demand. The risk of insufficient staff is, therefore, a function of this expression for loss. Since some demand for specialist type i also requires other types of specialist care of type i' at the same location due to co-morbidity, the total demand includes the product of the variable $X_{c,l,i'}$ with the co-morbidity factor, $m_{i,i'}$, summed over all other specialist types i' to account for the total demand at a cluster. To ensure that the clustering of specialty care follows the specified hierarchy (e.g. clustering low-priority specialist types does not cause high-priority demand to change location), the parameter $m_{i,i'} = 0$ for $i' > i$.

The second objective function (2) characterizes a system-wide penalty incurred from (i) paying for staffing, (ii) forcing patients to travel, and (iii) impacting patient familiarity and possible discontinuity with the clinics. The first term in (2) penalizes the hours of specialty type i staffed at each cluster i weighted by hourly cost $\sum_l Y_{l,i} \cdot h_{l,i}$.

The second term in (2) describes the linear travel penalty for transporting an hour of demand from its local clinic to any other clinic designated as its regional cluster. We express the travel penalty in terms of the average demand transported from clinic c as $\sum_{c=1}^{C}\sum_{l=1}^{C}(\sum_{i'}^{I} m_{i',i} \cdot \mu_{c,i'} \cdot t_{c,l,i'}) \cdot X_{c,l,i}$ for each specialty type i. This accounts for one type of demand following the re-direction of another type of demand due to co-morbidity. This formulation assumes that all transport costs are common to a region served by an existing clinic. Furthermore, we set the transportation penalty to zero if the clinic belongs to its own cluster, so by definition $t_{c,l} = 0$ when $c = l$.

The third term in the aggregated penalty function is a piecewise linear expression that accumulates penalty after a cluster's staffing level exceeds a threshold considered "large." Therefore, the penalty for having more than a threshold number of specialist-hours is $\sum_l [Y_{l,i} - v_{l,i}]^{+} \cdot s_{l,i}$.

Constraint (3) ensures that each staffing allocation in a cluster exceeds the expected weekly demand in that cluster. Constraint (4) ensures that each cluster l never exceeds a maximum number of hours that can be staffed. Every clinic is assigned to one and *only one* cluster as required by Constraints (5) and (8). Constraint (6) prevents clinics that provide care to a cluster from redirecting local demand to other clinics. Constraint (7) enforces the hierarchy of staffing specialist care, such that supporting specialist care with a low priority (high index i) is always staffed in clinics that support higher-priority specialist types. The last two constraints enforce binary values for cluster allocation, and non-negativity with respect to specialist hours.

The model provided in (1)−(9) determines an optimal policy that balances penalty versus the risk of demand for care exceeding staffed hours. We contrast this method of modeling the problem with an alternative method that treats the demand as deterministic. The decision variables are the same; the staffing levels ($Y_{l,i}$) and location of the staff ($X_{c,l,i}$).

To model the staffing problem deterministically, we replace the random variable of demand $D_{c,i}$ with its corresponding mean $\mu_{c,i}$. With this substitution, the risk of not meeting the expected demand in (1) is zero since Constraint (3) ensures the expected demand is satisfied. This simplifies the optimization problem, since minimizing the cost of

staffing ($\sum_{i=1}^{I}\sum_{l=1}^{C} Y_{l,i} \cdot h_{l,i}$), while ensuring that average demand is met (Constraint (3)), resulting in setting total staffing levels equal to the associated expected demand, such that

$$\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot \mu_{c,i} \cdot X_{c,l,i'}\right) = Y_{l,i} \ \forall l, i. \tag{10}$$

To locate staff, we then set $Y_{l,i}$ according to (10) and substitute the expression, in terms of $X_{c,l,i}$, into the penalty function. This results in the following minimization problem:

Minimize:

$$\begin{aligned}&\sum_{i=1}^{I}\left(\sum_{l=1}^{C}\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot \mu_{c,i} \cdot X_{c,l,i'}\right) \cdot h_{l,i}\right.\\ &+\sum_{c=1}^{C}\sum_{l=1}^{C}\left(\sum_{i'} m_{i',i} \cdot \mu_{c,i'} \cdot t_{c,l,i'}\right) \cdot X_{c,l,i}\\ &\left.+\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot \mu_{c,i} \cdot X_{c,l,i'}\right) - v_{l,i}\right]^{+} \cdot s_{l,i}\right)\end{aligned} \tag{11}$$

Subject to:

Constraints (4) – (8)

Note (3) is naturally satisfied by the substitution in (10) and the constraints are re-written in terms of the decision variables $X_{c,l,i}$. This provides staffing levels and cluster organization in a deterministic model. In Section 4.4, we compare this deterministic formulation with the multi-objective stochastic formulation given in (1)–(9). We demonstrate that the risk exposure using the deterministic model is greater than the risk associated with the solutions to the multi-objective stochastic formulation.

3.2. Constructing an efficient frontier

To demonstrate the tradeoffs between risk and penalty and provide optimal strategies to an administrator, a Pareto set of optimal points is generated. To construct this set, the multi-objective program listed in (1)–(9) is converted into a single objective optimization problem, with the second objective function (2) converted into a constraint with a variable right hand side $\tau_{penalty}$. This is referred to as the ϵ-method (Deb *et al.*, 2016) and creates an optimization problem with a single objective function and an additional constraint.

For large values of $\tau_{penalty}$, it is possible for the program to find unnecessarily large values of $Y_{c,l}$ that do not directly contribute to lowering the objective function. It is, therefore, important to include an exogenous term ($Y_{l,i}$ multiplied times a small constant ξ_i) to the objective function to ensure that excess staffing hours are not allocated. The single objective optimization problem is written as follows:

Minimize:

$$\begin{aligned}&\sum_{i=1}^{I}\left(w_i \cdot Risk_{\gamma_i}\left(\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}\right)\right.\\ &\left.+\xi_i \cdot \sum_{l=1}^{C} Y_{l,i}\right)\end{aligned} \tag{12}$$

Subject to:

$$\begin{aligned}&\sum_{i=1}^{I}\left(\sum_{l=1}^{C} Y_{l,i} \cdot h_{l,i} + \sum_{c=1}^{C}\sum_{l=1}^{C}\left(\sum_{i'} m_{i',i} \cdot \mu_{c,i'} \cdot t_{c,l,i'}\right) \cdot X_{c,l,i}\right.\\ &\left.+\sum_{l=1}^{C}\left(Y_{l,i} - v_{l,i}\right)^{+} \cdot s_{l,i}\right) \le \tau_{penalty}\end{aligned} \tag{13}$$

Constraints (3)–(9).

An efficient frontier for the multi-objective program in (1)–(9) can be created by solving a series of linear mixed-integer programs with different values of $\tau_{penalty}$.

We next find an approximation for the risk objective function.

3.3. Quantifying risk of staffing shortfall

The risk measure used to quantify under-staffing can take a variety of forms. In this article, we use CVaR as a measure of risk in the first objective function while ensuring that average weekly demand is met with Constraint (3). We estimate the risk function in (1) by using K-sampled values $d_{c,i,1}, \ldots, d_{c,i,K}$ from the demand $D_{c,i}$ distribution.

The most straightforward measure of risk is to measure the average demand that exceeds the number of staffed hours at any clinic cluster. Constraint (3) uses average demand expressed in terms of $\mu_{c,i}$; however, it could be expressed in terms of the K sample values

$$\begin{aligned}&\sum_{i=1}^{I}\left(w_i \cdot \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot d_{c,i,k} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}\right)\right.\\ &\left.+\xi_i \cdot \sum_{l=1}^{C} Y_{l,i}\right).\end{aligned}$$

However, the expected value of unsatisfied demand is a simple measure of risk that only represents central tendency and does not reflect the impact of the tail of the distribution.

An alternative to looking at expected staffing shortfall is using the VaR measure. The VaR at a given level γ_i represents the upper quantile of demand for specialist care exceeding staffed hours with probability γ_i. VaR (at level γ_i) provides a useful measure and is closely related to a chance constraint that specifies the probability of demand exceeding supply. For instance, if a decision maker required that demand not exceed supply with probability $1-\gamma'$, then this would be the equivalent of $VaR_{\gamma'} = 0$.

Table 2. *p*-Values, Anderson–Darling test.

Care type	Clinic 1	Clinic 2	Clinic 3	Clinic 4	Clinic 5	Clinic 6
Ambulatory care	4.12e − 26	1.11e − 05	3.92e − 04	3.92e − 06	6.41e − 12	4.38e − 12
Radiology	3.13e − 13	8.14e − 09	1.03e − 43	8.14e − 09	1.35e − 02	2.03e − 54
Lab only	9.02e − 61	2.61e − 49	1.11e − 46	8.76e − 50	5.70e − 57	5.80e − 58

Table 3. *p*-Values, Bonett–Seier test.

Care type	Clinic 1	Clinic 2	Clinic 3	Clinic 4	Clinic 5	Clinic 6
Ambulatory care	1.41e − 26	1.21e − 04	*7.21e−01*	*7.76e−01*	5.26e − 18	1.35e − 16
Radiology	5.72e − 08	*6.16e−01*	1.43e − 14	2.01e − 06	1.00e − 03	*6.84e−01*
Lab only	1.05e − 04	8.77e − 04	1.97e − 03	6.71e − 04	8.04e − 05	7.08e − 05

The objective function in *Risk* (1) can be written in terms of VaR as ($Risk_{\gamma_i} = VaR_{\gamma_i}$)

$$\sum_{i=1}^{I}\left(w_i \cdot VaR_{\gamma_i}\left(\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}\right) + \xi_i \cdot \sum_{l=1}^{C} Y_{l,i}\right).$$

This, combined with Constraints (3)–(9) and (13), results in a non-convex program. However, a heuristic approximation for the minimization of this quantity can be taken from solving successive lower bounds as outlined in Larsen *et al.* (2002). However, the "VaR" measure does not account for tail-end behavior past a specified threshold, and the measure can be not "sub-additive" and can miss opportunities where more consolidation could lead to less risk.

Lastly, the risk measure of "CVaR " at a given level γ_i ($CVaR_{\gamma_i}$) can be used to quantify an aversion to under-staffing. The CVaR risk measure can be seen as the expected amount of loss in the worst γ_i% of scenarios, or the average number of hours of specialist care in excess to the number of hours staffed in the worst γ_i% of scenarios.

The conditional value at risk has a number of qualities that make it preferable to the other measures of risk under consideration. CVaR has the advantage of being a convex and coherent measure of risk (Rockafellar, 2007). In our context, it ensures that the total system-wide risk of being understaffed is less than the sum of the risk at each cluster. Moreover, the use of CVaR allows the optimization problem to be sensitive to "tail-effects" that are an important component in staffing and clustering decisions that balance risk-pooling with minimizing the penalty objective function.

Additionally, CVaR is a natural upper bound for the VaR measure, making it a more conservative means to measure deviation from a given chance constraint. In fact, in many reliability applications "CVaR" is used as an equivalent to a "buffered chance constraint" (or buffered probability of failure in reliability contexts) (Rockafellar and Royset, 2010). This allows consideration of chance constraints to be accounted for without discounting large downsides that could result from rare scenarios on the tail-end of the distribution.

To estimate the $Risk_{\gamma_i}$ function in (1), we use a CVaR approximation method outlined by Rockafellar and Uryasev (1999) for the loss function, $\sum_{l}^{C}\left[\sum_{c}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}$. The CVaR term, $CVaR_{\gamma_i}\left(\sum_{l}^{C}\left[\sum_{c}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}\right)$, can be approximated using K demand samples, $d_{c,i,1}, \ldots, d_{c,i,K}$, from each $D_{c,i}$ and determining a threshold value, ϵ_i, that minimizes the average loss function values that exceed that threshold. The approximation from Rockafellar and Uryasev (1999) is given by

$$CVaR_{\gamma_i}\left(\sum_{l}^{C}\left[\sum_{c}^{C}\left(\sum_{i'}^{I} m_{i,i'} \cdot D_{c,i} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+}\right) \approx min_{\epsilon_i}\left(\epsilon_i + \frac{1}{K \cdot \gamma_i}\sum_{k=1}^{K}\left[\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot d_{c,i,k} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+} - \epsilon_i\right]^{+}\right). \quad (14)$$

As shown in Rockafellar and Uryasev (1999), the approximation in (14) is close to the $CVaR_{\gamma_i}$ of the loss function for sufficiently large values of K. The idea is that the value of ϵ_i that minimizes (14) approaches VaR_{γ_i} as K goes to infinity.

For an intuitive explanation of the approximation on the right-hand side of (14), consider two threshold values, $\epsilon_i^1, \epsilon_i^2$, such that $\epsilon_i^1 > VaR_{\gamma_i} > \epsilon_i^2$. Looking at the first term in the approximation, we see that $\epsilon_i^1 > \epsilon_i^2$. Conversely, the second term in the approximation has fewer terms above the threshold, ϵ_i^1, than the number of terms above the threshold ϵ_i^2. Therefore, the second term is smaller under ϵ_i^1 than under ϵ_i^2. There is a similar tradeoff between the first and second terms of the approximation when $\epsilon_i = VaR_{\gamma_i}$. The argument that the sum of both terms is *minimized* at the point where $\epsilon_i = VaR_{\gamma_i}$ depends on the convexity of $CVaR_{\gamma_i}$, as detailed in the proof of Theorem 1 from Rockafellar and Uryasev (1999).

We substitute the approximation in (14) for $Risk_{\gamma_i}$ in (12), since the decision variables ϵ_i can be selected independently for each i. This yields the following optimization problem with additional decision variables ϵ_i :

Minimize :

$$\sum_{i=1}^{I} w_i\left(\epsilon_i + \frac{1}{K \cdot \gamma_i}\sum_{k=1}^{K}\left[\sum_{l=1}^{C}\left[\sum_{c=1}^{C}\left(\sum_{i'} m_{i,i'} \cdot d_{c,i,k} \cdot X_{c,l,i'}\right) - Y_{l,i}\right]^{+} - \epsilon_i\right]^{+} + \xi_i \cdot \sum_{l=1}^{C} Y_{l,i}\right)$$

Subject To :

Constraints (3) − (9) and (13). (15)

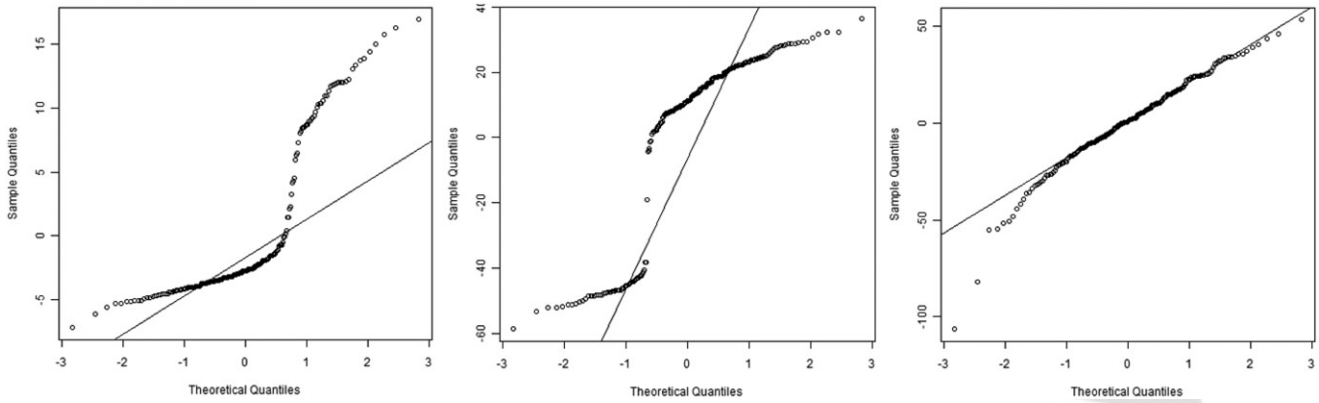

Figure 2. Three qq-plots that show the range of quantiles versus a normal behavior (straight line). Radiology in clinic 5 (on the right) follows the normal behavior for distributed weekly total RVUs, whereas radiology in clinic 6 (on the left) and lab-only in clinic 3 (in the middle) depart significantly at the lower and upper quantiles.

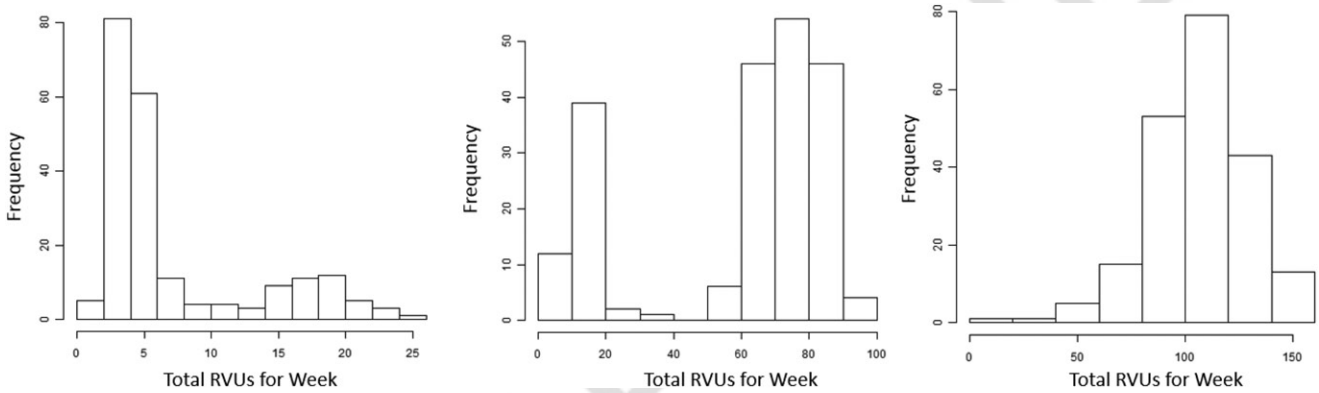

Figure 3. The histograms illustrating the distribution of RVUs for radiology in clinic 6 (on the left) and for lab-only in clinic 3 (middle) and radiology in clinic 5 (on the right). The leftmost graphs show non-normal bulges at the tail of the distribution whereas the rightmost graph shows demand consolidated around the mean (more normal behavior).

Since the objective function in (15) is piecewise linear and is monotonic relative to all positive-portion arguments, a change of variables can convert (15) into a linear equation.

4. Optimization on sample data

4.1. Evidence of non-normal demand

Using a dataset from Group Health, we demonstrate an initial motivation for a model to consider non-normal demand distributions. We use the Relative Value Unit (*RVU*) as a proxy for physician effort. It is a common way of measuring the time and effort of a physician for purposes of reimbursement. It is designed to account for the "time, technical skill, … and stress to provide a service" (Coberly, 2015).

The Group Health databases containing *RVU*s at various clinics provide insight into the fluctuation of patient demand. Initially tabulated by interaction, we aggregate the *RVU*s, by week for time period between January 2009 and December 2012, the *RVU*-score is totaled for each of six clinics and three categories of care (Ambulatory Visit, Lab Only Encounter, and Radiology). This creates an initial dataset of 208 values (each week for 4 years) that can be analyzed as a distribution.

Given the 208 weekly *RVU* totals, we apply some common statistical tests to demonstrate that the demand data does not follow a normal distribution. Using the null hypothesis that the observed data are taken from a normal distribution with unknown mean and variance, we generate *p*-values, the probability of observing sampled data under the null hypothesis, with the Anderson–Darling test and the Bonett–Seier test (Shapiro, 1990; Bonett and Seier, 2002). These values are listed in Tables 2 and 3. The null hypothesis of normality can be rejected in all 18 cases with 95% confidence using the Anderson-Darling test, and all but 4 of the 18 cases can be rejected with 95% confidence via the Bonett–Seier test (exceptions italicized).

We examine the behavior of three distributions (radiology at clinic 6, lab-only in clinic 3 and radiology in clinic 5) for illustration purposes (left, middle, and right in Figs. 2 and 3). The deviation from normal behavior of the three distributions can be visualized with a qq-plot (see Fig. 2) as well as in a histogram (see Fig. 3). The first two cases, radiology at clinic 6 and lab-only in clinic 3, show heavy tailed or skewed distributions that depart from normal behavior at the high and low quantiles. However, as seen in the third

Table 4. The staffing allocations for Scenario (I). Each of the percentages represents the distribution of the staffing across each of the seven clinics for each model, respectively. The bottom row contains the total weekly staffed hours for each specialist type.

	Oncology		Endocrinology		Behavioral care	
	Normal	Weibull	Normal	Weibull	Normal	Weibull
Clinic 1	36.99%	44.75%	36.09%	37.78%	29.11%	32.55%
Clinic 2	10.41%	–	10.65%	–	11.24%	11.34%
Clinic 3	8.89%	12.20%	10.65%	11.31%	11.24%	11.34%
Clinic 4	–	–	–	–	–	5.23%
Clinic 5	–	–	–	–	–	5.23%
Clinic 6	43.71%	43.05%	42.60%	45.25%	43.22%	29.07%
Clinic 7	–	–	–	5.66%	5.19%	5.23%
Total staffed hours in network	17,110	18,773	18,777	17,679	11,567	11,467

Table 5. The staffing allocations for Scenario 2. Each of the percentages represents the distribution of the staffing across each of the seven clinics for each model respectively. The bottom row contains the total weekly staffed hours for each specialist type.

	Oncology		Endocrinology		Behavioral Care	
	Normal	Weibull	Normal	Weibull	Normal	Weibull
Clinic 1	79.67%	77.71%	67.51%	78.77%	68.32%	78.30%
Clinic 2	–	–	–	–	–	–
Clinic 3	–	–	–	–	8.34%	–
Clinic 4	–	–	–	–	–	–
Clinic 5	–	–	–	–	–	–
Clinic 6	20.33%	22.29%	32.49%	21.23%	23.34%	21.70%
Clinic 7	–	–	–	–	–	–
Total staffed hours in network	31,355	27,977	24,625	24,572	15,591	15,187

case, radiology in clinic 5, some RVU distributions exhibit more normal behavior (rightmost graphs of Figs. 2 and 3).

As indicated by the data, the aggregated weekly demand does not seem to converge to a normal distribution, most likely because demand between weeks is not independent or is weakly dependent. This indicates that there is a diverse pattern behavior of demand such that a normal distribution might either over-estimate or under-estimate the tail effect of a distribution. Decision makers must be aware of cases where specialist demand exhibits heavy-tailed or skewed behavior and build these considerations into their optimization models.

Although there may be unobserved factors that account for the non-normality in these cases, the observations provide an initial motivation for testing cases involving non-normal distributions with *both* strong and weak tail-effects.

4.2. Seven clinic scenario with three specialties

To demonstrate the utility of our model, we examine a hypothetical scenario that emulates the behavior of a real-world network of healthcare providers. Although the information used to specify this scenario does not specifically fit the data discussed in Section 4.1, it is an example that demonstrates the optimization of specialist care allocation in a network, specifically demonstrating the utility of modeling a variety of demand distributions that can account for heavy-tailed and skewed demand distributions.

We compare two models that minimize the CVaR measure of risk based on two different demand distribution assumptions, normal and Weibull. For each of the model assumptions, both models will be fit to a common initial set of generated data points. For the initial generation of demand points, we specifically select the Weibull distribution as an example of a non-normal distribution that supports heavy-tail effects (similar to the behavior of the demand graphed in Section 4.1).

We investigate a hypothetical scenario with three specialist types: oncology, endocrinology, and behavior care (corresponding to morbidities of cancer, diabetes, and depression) across a seven-clinic network. We compare the allocation of these three types of specialists to the seven clinics under normal and Weibull demand distributions. The system parameter values are based, in part, on observations taken from literature and conjecture. In a real-world scenario, values would be determined through statistical analysis and an understanding of specific management priorities.

Starting with a seven clinic scenario similar to the example in Fig. 1, we model a network with two large clinics and five smaller clinics. From Group Health data, we observed that demand between larger and smaller clinics (based on RVU proxy) differs by a factor of 7. We also assume that the demand for specialist time differs between specialist types, since oncology care tends to be more time intensive than endocrinology, and endocrinology more intensive than behavioral care. Therefore, we scale the initial Weibull demand distribution for each clinic and specialist type such that large clinics have approximately seven times the demand of the smaller clinics and such that the demand for oncology, endocrinology, and behavioral demand differ by a factor of 1.5, respectively. At each clinic, the capacity is approximately twice the average demand with the exception

of the largest clinic (clinic 1), which we assume to be a large medical center with totally elastic staffing (arbitrarily large capacity). The threshold for considering a clinic "large" for the purpose of applying the discontinuity penalty is established as the capacity of a smaller clinic (for each specialty type).

We assume oncology and endocrinology to be more expensive to staff and they are therefore assigned a higher staffing cost. We understand the relative aversion to the risk of under-staffing as a function of patients' capacity to tolerate care disruptions. Based on the literature, we evaluate oncology patients to be less sensitive to disruptions in care than endocrinology and behavioral care (Reid *et al.*, 2005; Cho *et al.*, 2015). Travel penalties are multiplied by the physical distance between clinics, as shown on the map in Fig. 1. We understand that patients have a high aversion to traveling for endocrinology and behavioral care (Provost *et al.*, 2015), so that the penalties for these specialties are high. Similarly, patients are likely to prefer more familiarity for behavioral care than oncology and endocrinology (Provost *et al.*, 2015).

Finally, the matrix determining how a percentage of a given specialty should follow another specialty allocation is developed from an understanding of patients' co-morbidity and the prevalence of conditions in the specific population in question. In the absence of specific data, we use a baseline of 10% to represent the percent of each demand for certain type of specialist care that comes from patients already demanding another type of care (assuming that endocrinology follows oncology, and behavioral care follows both oncology and endocrinology).

With these considerations, we can populate the parameter values for both the normal and Weibull models. However, for a given situation, these parameters will have to be individually determined or weighed against the priorities of management. All parameter values are listed in Appendix A.

For the generation of the initial data, we model each clinic, for each specialty type i with a Weibull distribution $D_{c,i} \sim Weibull(\alpha_{c,i}, \beta_{c,i})$ with the shape parameter denoted as $\alpha_{c,i}$ and scale parameter as $\beta_{c,i}$. The scenario has two high-demand clinics (clinics 1 and 6) surrounded by five clinics (clinics 2, 3, 4, 5, and 7) with lower demand. We assign clinics 1 and 6 a shape parameter that will result in light-tails ($\alpha > 1$) and clinics 2, 3, 4, 5, and 7 a shape parameter that will result in heavy tails ($\alpha \leq 1$). See Table 6 in Appendix A.

Given the seven clinic demand distributions, we generate a set of 1000 sample data points for each specialty. We then fit each clinic's data with both a normal and Weibull distribution for comparison purposes. Based on the initial sample of 1000 points for each clinic we determine both Weibull $(\hat{\alpha}_{c,i}, \hat{\beta}_{c,i})$ and normal $(\hat{\mu}_{c,i}, \hat{\sigma}_{c,i})$ parameters using maximum-likelihood estimators (*MLEs*) (see Table 7 in Appendix A for fit parameters). We use each model to generate the $K = 1000$ sample points for the approximation of the objective function.

Based on all specified sample demands and system parameters, we examine the differences in recommended strategy when modeling the clinic demand as a normal distribution (which typically has a light tail) and modeling the clinic demand as Weibull distribution (which can have heavy tails based on the shape parameter). Given identical system parameters and constraints, these solutions are compared to demonstrate the importance of using a non-normal distribution for demand, when indicated by the data, for generating optimal staffing strategies.

4.3. Numerical results

To characterize the difference between demand modeling assumptions and optimal staffing, we solve the optimization problem with objective function specified in (15) and constraints in (3)–(9) and (13), based on each demand sample from the normal and Weibull models, respectively. For each comparison, we prepare an efficient frontier to demonstrate the tradeoff between the risk and penalty measures under each model, as well as a more detailed examination of optimal staffing recommendations for two specific scenarios corresponding to $\tau_{penalty}$ thresholds.

To illustrate the efficient frontier for both models, we select 10 discrete penalty threshold points, $\tau_{penalty} \in \{100000, 116000, 132000, 148000, 164000, 180000, 196000, 212000, 228000, 244000\}$. The values selected for $\tau_{penalty}$ range from the level where the program just becomes feasible ($\tau_{penalty} = 100000$) to where the optimal risk is no longer significantly changed ($\tau_{penalty} = 244000$). For each value of $\tau_{penalty}$, the optimization problem provides the minimum risk subject to the specified constraints.

Fig. 4 shows the tradeoff between risk and penalty for non-dominated solutions on the efficient frontiers both for the normal model (left) and the Weibull model (right). Here, the horizontal-axis represents the aggregated penalty whereas the vertical-axis represents the weighted linear combination of weekly demand hours that exceed supply (weighted CVaR in (15)). In each case, the graphed curves represent the tradeoff between the two model objectives. The number of clusters for each specialist type (oncology, endocrinology, and behavioral care) is given in the figure for each optimal solution.

Under the normal model, there is a large reduction in risk as $\tau_{penalty}$ increases beyond 100,000 with diminishing reductions in risk for further increase in $\tau_{penalty}$ beyond 180,000. Conversely, the Weibull model shows steeper drop-offs before $\tau_{penalty}$ exceeds 140,000 and does not flatten out as quickly. In both cases, we observe a tendency for fewer clusters to be used in the optimal solution of each specialist as the penalty allowance $\tau_{penalty}$ increases and the clinics are allowed to consolidate the risk into fewer high-capacity specialty centers.

To further compare the recommendations from the optimization, we explore the detailed solutions at two particular values of $\tau_{penalty}$ (as labeled in Fig. 4), (I) $\tau_{penalty} = 116000$, (II) $\tau_{penalty} = 228000$ which represent optimal specialist staffing arrangements under low $\tau_{penalty}$ constraint and high $\tau_{penalty}$, respectively. The Pareto-optimal strategies generated by the optimization change depending on the distribution used to fit the sample demand data, as shown in Figs. 5 and

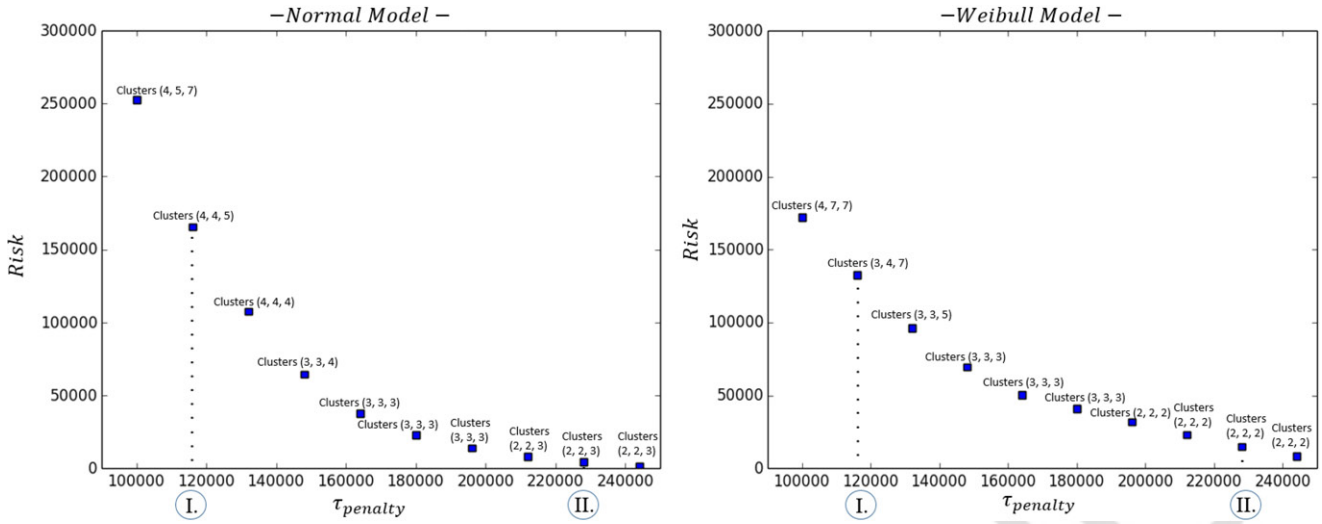

Figure 4. The constructed efficient frontier (non-dominated solutions) for the optimization model where demand is modeled with a normal distribution (left panel) and a Weibull distribution (right panel). Each data point is labeled with the number of clusters for each specialist type in the final optimized solution (in order: oncology, endocrinology, and behavioral care).

6. A detailed breakdown of the optimal staffing levels at the clinics for each model is given in Tables 4 and 5.

Generally, both models show some similarity in their corresponding optimal solutions for each scenario. In Scenario (I), both Weibull and normal models generally recommend a much more distributed strategy due to the low penalty value. In contrast, for Scenario (II), both models recommend a highly centralized strategy (since the larger penalty allowance allows for more risk pooling). This illustrates the trade-off between risk and penalty representing staffing cost, patient travel time, and familiarity penalty. The similarity of strategies between the two scenarios is strongest in oncology care, where both Weibull and normal models recommend identical clustering in all but one case.

However, there are distinct differences between the models' optimal staffing recommendations. In Scenario (I), the normal model recommends some consolidation of behavioral care, whereas the Weibull model recommends each clinic support its own behavior care while at the same time creating slightly centralized provision of the oncology care. In Scenario (II), the Weibull model recommends a more consolidated strategy for behavioral care but, in fact, generally staffs fewer specialist hours overall.

The difference in these optimal strategies may be due to under-estimating the risk value with the normal distribution so that the benefits of risk-pooling are not valued in the same way as with the Weibull distribution. This again can be seen in the more consistent clustering of oncology care where risk aversion is low, to the larger differences between the models in recommended behavioral care clustering where risk aversion is higher. The sensitivity to a risk of care disruption makes an optimal solution more dependent on the model for demand distribution.

The change in optimal clustering arrangement between the data fit with a normal model and Weibull model demonstrates the importance of being able to accurately fit a distribution that represents the tail-behavior of demand. If heavy or truncated tail behavior in demand is noticed, it may be important for a decision maker to use a model that does not assume normally distributed demand in order to reduce the risk that patients will experience care disruptions when seeking specialist care.

4.4. Comparing the multi-objective stochastic model to a deterministic model

The solutions generated by the multi-objective stochastic model are compared to a solution determined by the alternative deterministic model described in Section 3.1. We compare the solutions in terms of the risk exposure they produce.

Using the initial sample demand points, we examine the approximate CVaR risk under each policy solution (note these values are different than the estimated risk levels generated from solving the multi-objective problem). Here, we approximate the risk of staffing shortfall by averaging the demand in excess of staffed hours at the $\delta = 0.05$ quantile level under each policy solution. Fig. 7 illustrates the approximated CVaR risk for the policy solution from the deterministic model, and from Scenarios (I) and (II) from the Weibull model (Figs. 4–6).

Based on the comparison, we see that the deterministic model generates higher risk than the solutions generated with the multi-objective stochastic model. This generally suggests that the deterministic model prioritizes a low-penalty function at the cost of increased risk of deferred care and long wait times.

The solution provided is generally more distributed (i.e. more clusters), which reduces patient travel time and reduces any penalty for large clinics (the staffing expense is as low as possible given the average demand constraint in (3)). Thus, the use of the deterministic model results in risk-insensitive solutions that expose patients to further risk of not having sufficient staffing to meet their needs in a timely manner.

Therefore, the benefit of the multi-objective stochastic model is two-fold: first, it captures the interaction between

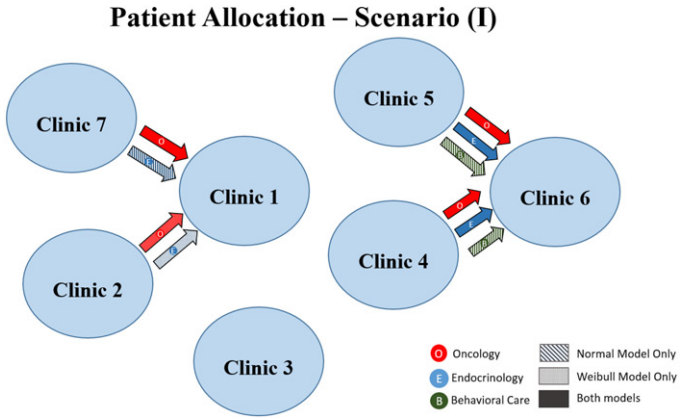

Figure 5. (I) $\tau_{penalty} = 116000$. The lined-arrows show re-direction of patients under the normal model, the dotted arrows show the redirection of patients recommended by the Weibull model, and the solid colors show the re-direction under both models. Colors and labels distinguish the different specialist types being directed.

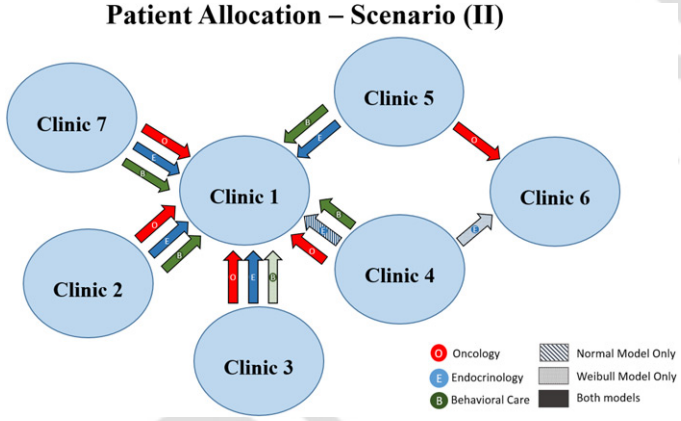

Figure 6. (II) $\tau_{penalty} = 228000$. The lined-arrows show re-direction of patients under the normal model, the dotted arrows show the redirection of patients recommended by the Weibull model, and the solid colors show the re-direction under both models. Colors and labels distinguish the different specialist types being directed.

staffing level and staff location, which results in an awareness of risk exposure, and, second, it allows a decision maker to understand the tradeoffs between risk and penalty formulations by weighing penalty cost against exposure to the risk of staffing shortfall.

5. Discussion

This article provided an optimization model for allocating specialty staff across a set of geographically distributed medical centers based on minimizing the risk of care disruption and the average system penalty costs. Due to the impact of centralized staff becoming increasingly necessary (Rais and Viana, 2011), a tool designed to quantify the risks of patient-centered staffing strategies will be useful to administrators. Moreover, our model will allow an administrator to measure the risk and penalty tradeoff implicit in status-quo staffing strategies and compare the results to an optimized allocation. Combined with domain knowledge, the information furnished by these solutions could lead to more efficient staffing with lower risk, less wait time, and minimized inconvenience for patients.

One benefit of our model is its ability to address non-normal distributions and consider the critical impact of heavy-tailed effects on risk and penalty. As demonstrated in Section 4.3, accounting for tail effects can be potentially important when generating an optimal staffing strategy. Therefore, a model that minimizes risk based on a general distribution could provide an advantage over others that use an assumption of normally distributed demand. Yet another benefit to our model is the ability to generate solutions along the efficient frontier to provide insight into balancing risk with penalty costs. Further examination shows that the policy solutions supplied by the multi-objective stochastic model result in lower risk policy recommendations than the alternative of solving the problem deterministically. The use of the multi-objective stochastic model allows a decision maker to directly observe the tradeoff between penalty and risk and, therefore, directly address risk aversion, allowing for greater consideration of patient needs to play a part in the decision making process.

Depending on the specific situation, administrators may need to modify the assumptions used in our model. In fact, when specific thresholds for total staffing, patient familiarity, or total travel exist, linear constraints could be added without significantly modifying the mixed-integer linear program formulation. For instance, many administrators consider specialist staffed hours to be non-flexible and might also

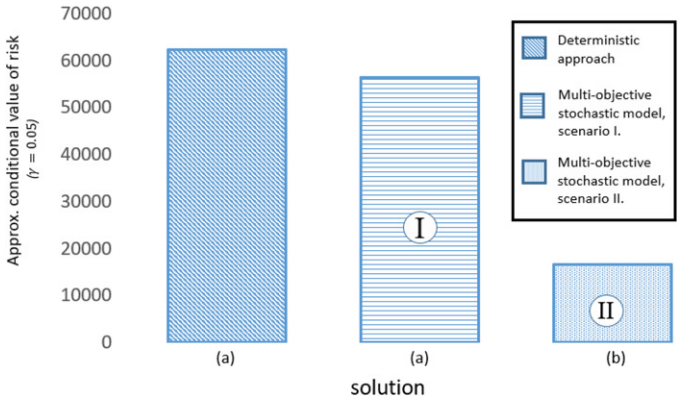

Figure 7. A comparison of the risk of demand hours exceeding staffed hours in three solutions using the (a) deterministic model, (b) multi-objective stochastic model, Scenario (I), and (c) multi-objective stochastic model, Scenario (II). Clusters (*x*, *y*, *z*) represent the number of clusters for (oncology, endocrinology, behavioral care) in each solution.

have limits on the total average distance traveled by patients in a given week. This addition may shed light on the particular constraints that force clustering and isolation of clinics. In either of these instances simple linear constraints could be added to the general optimization problem without altering the solution approach demonstrated in this article.

The proposed staffing model has several areas where extensions could provide results that are more practical and accurate in accounting for the risk and penalty that result from staffing medical specialists. In a future model, a clinic might have limits both on the total amount of staffing supported across all clinics as well as additional penalties for having a large number of specialists centered at the same clinic. Moreover, an updated model might also look into how the staffing of medical professionals in clinics might change when demand for different types of care are modeled with joint distributions.

Funding

Q1 This research was supported by National Science Foundation.

References

Abri, S. M., WestJr, D. J. , and Spinelli, R. J. (2006) Managing overutilization, quality of care, and sustainable health care outcomes in Oman. *The Health Care Manager*, **25**(4), 348–355.

Alexander, G. J., and Baptista, A. M. (2004) A comparison of var and cvar constraints on portfolio selection with the mean-variance model. *Management Science*, **50**(9), 1261–1273.

Baray, J., and Cliquet, G. (2013) Optimizing locations through a maximum covering/p-median hierarchical model: maternity hospitals in France. *Journal of Business Research*, **66**(1), 127–132.

Benneyan, J. C., Musdal, H., Ceyhan, M. E., Shiner, B., and Watts, B. V. (2012) Specialty care single and multi-period location allocation models within the veterans health administration. *Socio-Economic Planning Sciences*, **46**(2), 136–148.

Birge, J. R., and Louveaux, F. (1997) *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY.

Bodenheimer, T., and Pham, H. H. (2010) Primary care: Current problems and proposed solutions. *Health Affairs (Project Hope)*, **29**(5), 799–805.

Bonett, D. G., and Seier, E. (2002) A test of normality with high uniform power. *Computational Statistics and Data Analysis*, **40**(3), 435–445.

Calvo, A. B., and Marks, D. H. (1973) Location of health care facilities: An analytical approach. *Socio-Economic Planning Sciences*, **7**(5), 407–422.

Chen, G., Daskin, M. S., Shen, Z.-j. M., and Uryasev, S. (2006) The α-reliable mean-excess regret model for stochastic facility location modeling. *Naval Research Logistics*, **53**, 617–626.

Cho, K. H., Lee, S. G., Jun, B., Jung, B.-Y., Kim, J.-H., and Park, E.-C. (2015) Effects of continuity of care on hospital admission in patients with type 2 diabetes: Analysis of nationwide insurance data. *BMC Health Services Research*, **15**(1), 1.

Coberly, S. N. H. P. F. (2015) Relative Value Units. Tech. rep., George Washington University, Washington, DC.

Cohen, M. A., and Lee, H. L. (1985) The determinants of spatial distribution of hospital utilization in a region. *Medical Care*, **23**(1), 27–38.

Cook, N., Hicks, L., O'Malley, A., and Keegan, T. (2007) Access to specialty care and medical services in community health centers. *Health Affairs (Millwood)*, **26**(5), 1459–1468.

Daskin, M. S., and Dean, L. K. (2004) Location of health care facilities. In *Operations Research and Health Care: A Handbook of Methods and Applications*, vol. 70 of International Series in Operations Research and Management Science. Springer US, Boston, MA, pp. 43–76.

Deb, K., Sindhya, K., and Hakanen, J. (2016) Multi-objective optimization. In *Decision Sciences: Theory and Practice*. CRC Press, London, pp. 145–184.

Exworthy, M., and Peckham, S. (2006) Access, choice and travel: Implications for health policy. *Social Policy & Administration*, **40**(3), 267–287.

Gaynor, M., and Anderson, G. F. (1995) Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds. *Journal of Health Economics*, **14**(3), 291–317.

Griffin, P. M., Scherrer, C. R., and Swann, J. L. (2008) Optimization of community health center locations and service offerings with statistical need estimation. *IIE Transactions*, **40**(9), 880–892.

Halm, E. A., Lee, C., and Chassin, M. R. (2002) Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine*, **137**(6), 152.

Huckman, R. S., and Pisano, G. P. (2006) The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, **52**, 473–488.

Jack, E. P., and Powers, T. L. (2009) A review and synthesis of demand management, capacity management and performance in healthcare services. *International Journal of Management Reviews*, **11**(2), 149–174.

Jordan, H., Roderick, P., Martin, D., and Barnett, S. (2004) Distance, rurality and the need for care: Access to health services in south west England. *International Journal of Health Geographics*, **3**(1), 21.

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, **12**(2), 479–502.

Kohli, S., Sahln, K., Sivertun, A., Lfman, O., Trell, E., and Wigertz, O. (1995) Distance from the primary health center: A GIS method to study geographical access to health care. *Journal of Medical Systems*, **19**(6), 425–436.

Larsen, N., Mausser, H., and Uryasev, S. (2002) Algorithms for optimization of value-at-risk. In *Financial Engineering, E-commerce and Supply Chain*, P. M. Pardalos and V. K. Tsitsiringos, Eds. Springer US, Boston, MA, pp. 19–46.

Li, L., and Benton, W. (2003) Hospital capacity management decisions: Emphasis on cost control and quality enhancement. *European Journal of Operational Research*, **146**(3), 596–614.

Luft, H. S., and Crane, S. (1980) Regionalization of services within a multihospital health maintenance organization. *Health Services Research*, **15**(3), 231–247.

Maass, K., Liu, B., Daskin, M., Duck, M., Wang, Z., Mwenesi, R., and Schapiro, H. (2017) Incorporating nurse absenteeism into staffing

with demand uncertainty. *Health Care Management Science*, **20**(1), 141–155.

Mahar, S., Bretthauer, K. M., and Salzarulo, P. A. (2011) Locating specialized service capacity in a multi-hospital network. *European Journal of Operational Research*, **212**(3), 596–605.

Mihaylova, B., Briggs, A., O' Hagan, A., and Thompson, S. G. (2011) Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, **20**(8), 897–916.

Oliveira, M., and Bevan, G. (2006) Modelling the redistribution of hospital supply to achieve equity taking account of patient's behaviour. *Health Care Management Science*, **9**(1), 19–30.

Parker, J., and DeLay, D. (2008) The future of the healthcare supply chain: Suppliers wield considerable power, but healthcare organizations can benefit from virtual centralization of the supply chain.(feature story). *Healthcare Financial Management*, **62**(4), 66–69.

Payne, S., Jarrett, N., and Jeffs, D. (2000) The impact of travel on cancer patients experiences of treatment: a literature review. *European Journal of Cancer Care*, **9**(4), 197–203.

Pfeiffer, P. N., Glass, J., Austin, K., Valenstein, M., McCarthy, J. F., and Zivin, K. (2011) Impact of distance and facility of initial diagnosis on depression treatment.(quality and outcomes)(report). *Health Services Research*, **46**(3), 768–786.

Provost, S., Prez, J., Pineault, R., Borgs Da Silva, R., and Tousignant, P. (2015) An algorithm using administrative data to identify patient attachment to a family physician. *International Journal of Family Medicine*, **2015**, 1–11.

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

Reid, R. J., Scholes, D., Grothaus, L., Truelove, Y., Fishman, P., Mcclure, J., Grafton, J., and Thompson, R. S. (2005) Is provider continuity associated with chlamydia screening for adolescent and young adult women? *Preventive Medicine*, **41**(5), 865–872.

Rockafellar, R., and Royset, J. (2010) On buffered failure probability in design and optimization of structures. *Reliability Engineering and System Safety*, **95**(5), 499–510.

Rockafellar, R. T. (2007) Coherent approaches to risk in optimization under uncertainty. In *OR Tools and Applications: Glimpses of Future Technologies*. Informs, Hanover, MD, pp. 38–61.

Rockafellar, R. T., and Uryasev, S. (1999) Optimization of conditional value-at-risk. *Journal of Risk*, 1–26.

Rohleder, T., Bischak, D., and Baskin, L. (2007) Modeling patient service centers with simulation and system dynamics. *Health Care Management Science*, **10**(1), 1–12.

Ruth, R. J. A. (1981). A mixed integer programming model for regional planning of a hospital inpatient service. *Management Science*, **27**(5), 521–533.

Saha, S., Beach, M. C., and Cooper, L. A. (2008) Patient centeredness, cultural competence and healthcare quality. *Journal of the National Medical Association*, **100**(11), 1275–1285.

Shapiro, A. (2009) *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM series on optimization; 9. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Shapiro, S. S. (1990) *How to Test Normality and other Distributional Assumptions* (Vol. **3**). American Society for Quality Control, Milwaukee, WI.

Shen, S., and Chen, Z. (2013) Optimization models for differentiating quality of service levels in probabilistic network capacity design problems. *Transportation Research Part B*, **58**(C), 71–91.

Shuman, L. J., Hardwick, C. P., and Huber, G. A. (1973) Location of ambulatory care centers in a metropolitan area. *Health Services Research*, **8**(2), 37–56.

Sikka, V., Luke, R. D., and Ozcan, Y. A. (2009) The efficiency of hospital-based clusters: evaluating system performance using data envelopment analysis (report). *Health Care Management Review*, **34**(3), 251–261.

Smith-Daniels, V., Schweikhart, S. B., and Smith-Daniels, D. E. (1988) Capacity management in health care services: Review and future research directions. *Decision Sciences*, **19**(4), 889–919.

Stummer, C., Doerner, K., Focke, A., and Heidenberger, K. (2004) Determining location and size of medical departments in a hospital network: A multiobjective decision support approach (author abstract). *Health Care Management Science*, **7**(1), 63–71.

Trinh, H. Q., Begun, J. W., and Luke, R. D. (2014) Service duplication within urban hospital clusters (abstract). *Health Care Management Review*, **39**(1), 251–261.

Wang, W., and Ahmed, S. (2008) Sample average approximation of risk-averse stochastic programs, *Operations Research Letters*, **36**(5), 515–519.

Woods, M. D., Kirk, D., Agarwal, S., Annandale, E., Arthur, T., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., and Smith, L. (2005) Vulnerable groups and access to health care: a critical interpretive review. National Coordinating Centre NHS Service Delivery Organ RD (NCCSDO). (accessed May 27, 2012).

Yu, X., Sun, H., and Chen, G. (2011) The optimal portfolio model based on mean-cvar. *Journal of Mathematical Finance*, **01**(03), 132–134.

Zheng, Q. P., Shen, S., and Shi, Y. (2015) Loss-constrained minimum cost flow under arc failure uncertainty with applications in risk-aware kidney exchange. *IIE Transactions*, **47**(9), 961–977.

Zhu, S., and Fukushima, M. (2009) Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155–1168.

Appendix A: Parameter values for sample run

Listed below are the selected parameter values used to create the Weibull and normal models in Section 4.2. The results are labeled by clinics such that each cell contains the values in order of specialist type (ordered: oncology, endocrinology, and behavioral care).

Table 6. Weibull parameters.

Clinic	Shape: $\alpha_{c,i}$	Scale: $\beta_{c,i}$
1	5, 5, 5	5000, 4000, 3000
2	1, 1, 1	1500, 1000, 650
3	1, 1, 1	1500, 1000, 650
4	0.5, 0.5, 0.5	750, 500, 300
5	0.5, 0.5, 0.5	750, 500, 300
6	5, 5, 5	5000, 4000, 2500
7	0.5, 0.5, 0.5	750, 500, 300

Table 7. Fit parameters for normal and Weibull models.

Clinic	Modeled normal mean: $\hat{\mu}_{c,i}$	Modeled normal standard deviation: $\hat{\sigma}_{c,i}$	Modeled Weibull shape: $\hat{\alpha}_{c,i}$	Modeled Weibull scale: $\hat{\beta}_{c,i}$
1	4553, 3672, 2741	1056, 854, 644	5.19, 4.99, 4.95	5048, 4007, 3008
2	1611, 973, 640	1669, 943, 629	1.02, 0.99, 1.00	989.32, 655, 636
3	1512, 1003, 619	1479, 1013, 596	1.01, 1.01, 1.01	1577, 1020, 671
4	1600, 876, 703	3866, 1852, 1723	0.51, 0.53, 0.50	759.08, 539, 305
5	1529, 1082, 595	2996, 2683, 1287	0.47, 0.50, 0.494	715, 458, 328
6	4578, 3664, 2282	1062, 874, 535	4.89, 4.98, 5.07	5032, 3978, 2507
7	1594, 938, 632	3660, 2031, 1526	0.50, 0.52, 0.51	706, 563, 318

Table 8. System parameters.

γ_i	K	ξ	w_i
0.05	1000	0.0001	1, 2, 2

Table 9. Co-morbidity matrix $m_{i,i'}$

	Specialist type 1	Specialist type 2	Specialist type 3
Specialist type 1	1	0	0
Specialist type 2	0.1	0.9	0
Specialist type 3	0.1	0.1	0.8

Table 10. Transport penalty: $t_{c,l}$

	Clinic 1	Clinic 2	Clinic 3	Clinic 4	Clinic 5	Clinic 6	Clinic 7
Clinic 1	0, 0, 0	0.5, 1, 1	3, 6, 6	3, 6, 6	3.5, 7, 7	3.5, 7, 7	2, 4, 4
Clinic 2	0.5, 1, 1	0, 0, 0	3, 6, 6	2.5, 5, 5	3, 6, 6	3, 6, 6	4, 8, 8
Clinic 3	3, 6, 6	3, 6, 6	0, 0, 0	6, 12, 12	6.5, 13, 13,	7, 14, 14,	8, 16, 16,
Clinic 4	3, 6, 6	2.5, 5, 5	6, 12, 12	0, 0, 0	3, 6, 6	3, 6, 6	2.5, 5, 10
Clinic 5	3.5, 7, 7	3, 6, 6	6.5, 13, 13	1.5, 3, 3	0, 0, 0	1.5, 3, 3	5.5, 11, 11
Clinic 6	3.5, 7, 7	3, 6, 6	7, 14, 14	1.5, 3, 3	1.5, 3, 3	0, 0, 0	6.5, 11, 11
Clinic 7	2, 4, 4	4, 8, 8	8, 16, 16	5, 10, 10	5.5, 11, 11	5.5, 11, 11	0, 0, 0

Table 11. Clinic data.

Clinic	Capacity: $b_{l,i}$	Cost per staff: $h_{l,i}$	Discontinuity penalty rate: $s_{l,i}$	Discontinuity threshold: $v_{l,i}$
1	300,000, 300,000, 300,000	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
2	3000, 2000, 1300	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
3	3000, 2000 1300	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
4	1500, 1000, 600	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
5	1500, 1000, 600	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
6	10,000, 8000, 5000	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300
7	1500, 1000, 600	2, 1, 0.75	0.5, 0.3, 1	3000, 2000, 1300

Values generated for $(d_{c,i,1}, d_{c,i,2}, \ldots, d_{c,i,K})$ can be provided upon request.