Staff Planning for Hospitals with Cost Estimation and Optimization

Sandeep Rath
Kenan-Flagler Business School, University of North Carolina at Chapel Hill
Kumar Rajaram
UCLA Anderson School of Management, Los Angeles CA 90095

We consider the anesthesiologist staff planning problem for operating services departments in large multi-specialty hospitals. In this problem, the planner makes monthly and daily decisions to minimize total costs. The monthly decisions include deciding how many anesthesiologists should be on regular duty and how many should be on-call for each day of the month and for each specialty. The daily decisions involve determining how many on-call anesthesiologists to actually use in the surgical schedule for the next day. Total costs comprise of explicit and implicit costs. Explicit costs include the costs of calling an anesthesiologist from call, and overtime costs, and are specified by the organization. Implicit costs are the costs of keeping but not calling an anesthesiologist on-call and under-utilizing an anesthesiologist, and these have to be deduced from past decisions. We model the staff planning problem as a two-stage integer stochastic dynamic program, develop structural properties of this model and use them in a sample average approximation algorithm constructed to solve this problem. We also develop a procedure to estimate the implicit costs, which are included in this model. Using data from the operating services department at the UCLA Ronald Reagan Medical Center, our model shows the potential to reduce overall costs by 13%. We provide managerial insights related to the relative scale of these costs, hiring decisions by specialty, sensitivity to cost parameters, and improvements in prediction of booked time durations.

1. Introduction

Healthcare expenditure in the US is expected to rise to 20% of GDP by 2020 (Keehan et al. 2017). There is evidence to suggest that a significant portion of this expenditure is wasted due to operational inefficiencies at healthcare sites such as hospitals, which constitute around 32% of healthcare expenditure in the US (Smith et al. 2012). In hospitals, the total labor expenditure can exceed 50% of operating costs and can be up to 90% of variable costs (Healthcare Insights 2014). Thus, efficient deployment of labor becomes one of the primary methods of cost control at hospitals.

Managing labor at hospitals is challenging because of the uncertainty in demand for services, the specialized skill set of staff and since tactics such as production smoothing cannot be employed. Hospitals mitigate the effects of these challenges by using staffing resources that can be made flexible in volume by calling additional employees, use of floating resources, and through overtime (Kesavan et al. 2014). Such volume flexibility can help reduce cost at hospitals by reacting to changes as information about the future workload becomes available (Bard and Purnomo 2005).

Volume flexibility in hospitals has been used in staff planning for nurses and physicians (Brunner et al. 2009).

Overtime is a key feature in achieving volume flexibility. However, excessive overtime of clinical staff has been associated with lower patient safety (Rogers et al. 2004), higher employee burnout (Stimpfel et al. 2012), and deteriorating employee health (Trinkoff et al. 2006). Thus, to reduce reliance on overtime, staff planners often use additional employees who can be called on a short notice. The use of this contingency labor supply reduces the number of overtime hours. However, depending on the staffing policy, this may give rise to additional administrative costs. These costs consist of both explicit and implicit costs. Explicit costs represent the actual monetary payment made and recorded for an activity. Such costs could include overtime compensation and extra payments made to staff who work on a short notice. In contrast, an implicit cost is not recorded but implied. This could include the opportunity cost to the organization associated with staff idle time and the inconvenience to employees due to changing their schedule at a short notice.

Traditionally, staff planning at hospitals has been a manual process. While evidence suggests the use of analytic, data-driven, model based systems would be beneficial from a cost perspective (Healthcare Insights 2014), implementing such systems for labor scheduling has been challenging. There have been examples of automated staff planning systems that have not been successful at large retail organizations like Starbucks (Kantor 2014, 2015). The principal challenge in implementation of model based staff planning systems is minimizing overall costs by incorporating the explicit and implicit human costs of the employees being scheduled. Not incorporating all the human costs would likely lead to failure in acceptance and implementation of these systems (Bernstein et al. 2014)

In this paper we provide an approach to estimate the implicit costs in staff planning and subsequently, we use both explicit and implicit costs in an optimization model for anesthesiologist staff planning at the UCLA Ronald Reagan Medical Center (RRMC).

1.1. Problem Description

The UCLA RRMC is a large multi-specialty hospital which consistently ranks amongst the best five hospitals in the United States[1]. The operating services department of the UCLA RRMC is responsible for staffing anesthesiologist physicians to surgical services at the hospital. The focus of our work is the staff planning of physician anesthesiologists at this department of the UCLA RRMC.

The operating services department manages the surgery suite at the UCLA RRMC. Surgeons across multiple specialties in this hospital and from other hospitals perform around 27,000 surgeries

[1] http://health.usnews.com/health-care/best-hospitals/articles/best-hospitals-honor-roll-and-overview

annually across 2,700 unique surgery types. The anesthesia services provided for these surgeries belong to 4 service types: Cardio-Thoracic, General, Neuro, and Pediatric. The staff planning for anesthesiologist consists of two stages: monthly, and daily decisions. The details of these decisions are given below.

- **Monthly decisions:** On the 20^{th} of each month, depending on the teaching and vacation schedule of anesthesiologists, the availability of anesthesiologists for each day of the upcoming month is known. Once the anesthesiologist has provided their availability, they can be scheduled across all these days. Based on this availability, and the historical data of surgical work load, the staffing plan for each service for each day of the next month is prepared. This plan consists of dividing the anesthesiologists available on each day of the next month into two groups: those who would be available on regular duty, and those on a reserve list, called the *on-call* consideration list. Anesthesiologists on the on-call consideration list are informed the day before the surgery if their services are required the next day. In this case, they are paid an additional $1000 for the entire day. However, if they are not required, they are not paid this additional amount. Thus, being on the on-call consideration list and not being called is not desirable for the employees. Thus, the planner manages the number of employees they have on the on-call consideration list so that this does not occur frequently.
- **Daily decisions** The day before the surgery the total number of elective procedures that will be performed the next day and their booked hours is finalized. Based on this information, a certain number of anesthesiologist of each specialty from the on-call consideration list are informed that they would be working the next day. This determines the total available work hours. When the actual surgical hours are realized, the costs of overtime or idle time are realized.

The staff planner has to balance four costs when making the monthly and daily decisions involved in the staffing plan. These include:

1. The explicit cost of calling anesthesiologists from the on-call consideration list. This is the additional payment made to the anesthesiologists for coming in at a short notice. At UCLA RRMC this is $1000 per day.
2. The implicit cost of having anesthesiologists on the on-call consideration list but not calling them. This is the inconvenience cost of keeping an anesthesiologist on hold for a day and not compensating them.
3. On the day of surgery, after the total number of surgical hours have realized, if the total demand for anesthesia exceeds the total standard hours across all working anesthesiologists, overtime would need to be paid. At UCLA RRMC this overtime payment is around $180/hr.

4. If the total number of work hours of available anesthesiologists is greater than the total realized hours of surgery, there will be idle time. The operating services department does not prefer idle time and thus, there is an implicit cost of idle time.

In Table 1 we present the summary statistics of the number of anesthesiologists in regular duty, in the on-call consideration list, and who actually get called. This table shows that on an average there are 17.48 anesthesiologists working on regular duty, 6.89 are on the on-call list out of which 2.77 are actually called. Furthermore, there is considerable variation in staffing levels across specialties. This is primarily due to the demand characteristics of the specialties.

Insert Table 1 here

In 2014, the UCLA RRMC instituted an electronic health system[2]. The management at the operating services department was keen on using the data from this system to develop an analytical model based approach to staff planning which incorporated all the relevant costs. An implementation of such an analytical model to address staff planning could face similar challenges as described in Kantor (2014), Kantor (2015) and Bernstein et al. (2014), if implicit human costs of the staff being scheduled are not incorporated. Therefore, we take a two part approach to staff planning at this hospital. In the first part, we model the staff planning as a two-stage integer stochastic dynamic program. The first stage captures the monthly decisions, while the second stage includes the daily decisions involved in staff planning. For given cost parameters, we develop an algorithm to solve this model to provide the monthly and daily anesthesiologist staffing plan across each specialty. In the second part, we develop a procedure to estimate the implicit costs. These include the inconvenience costs of scheduling anesthesiologists on the on-call consideration list but not calling them, and the implicit cost of idle time. Subsequently, we use these estimated costs to demonstrate the total cost savings from using the optimization model.

1.2. Literature Review

The staff planning problem considered in this paper is related to three streams of literature. The first is in staff planning for services, particularly for operating rooms. The second stream is on two-stage stochastic dynamic programming models. The third is associated with estimation of operational parameters.

There have been several papers which model the stages of staff planning at service organizations as a dynamic optimization problem. Wild and Schneewei (1993) provide a model for staff planning for the long term, medium term and short term planning when volume flexibility in terms of contingent workers are available. Pinker and Larson (2003) provide a model for flexible workforce management in environments with uncertainty in the demand for labor. With respect to staff

[2] http://careconnect.uclahealth.org/about-careconnect

planning at hospitals, Dexter et al. (2005) provide a framework for tactical decision making for allocating operating room time approximately a year in advance. The decisions that are a part of this time-frame is hiring of additional staff and building new operating rooms. He et al. (2012) analyze decision making for nurse staffing as more information is available about the workload on the day of the surgery. Through numerical analysis they identify that deferring staffing decisions to a point when procedure type information is available could help hospitals save up to 49% of staffing costs. While hospitals would like to defer staffing decisions as late as possible, this often tends to staff not having their schedules finalized until a little before the day of the surgery. This uncertainty in schedules is not desirable from a staff perspective. Thus, the UCLA RRMC, like several other service organization mitigates this problem by using a base level of staff who know they will definitely be required at a given day, and a reserve list (on-call) who will know if they need to come in only the day before. McIntosh et al. (2006) state that this refinement of service specific staffing, months before the day of the surgery has a high degree of impact on staff satisfaction at hospitals. Xie and Zenios (2015) analyze the nurse staff planning problem within a time frame of a few months and propose dynamic staffing policy with adjustments to staffing levels as information on different types of surgeries arrives sequentially. They find that a threshold policy with two adjustment levels is optimal.

The staff planning problem at the UCLA RRMC is a two-stage, integer stochastic dynamic program. When we remove the integrality requirement, this problem reduces to a two-stage stochastic dynamic program. Such problems have been extensively studied (Birge 1985). When applied in the retail context, this is known as a two-stage news-vendor problem. Gurnani and Tang (1999) characterize the optimal solution for this problem at a retailer who has two instants to order a seasonal product. Fisher et al. (2001) propose a heuristic solution to solve the two-stage news-vendor problem in an application at a catalog retailer. Recently, such two-stage models have also been used in agro-business (Bansal and Nagarajan 2017). In contrast, integrality requirements in our problem are essential since we consider staff planning and as shown in Table 1, the average number of anesthesiologists deployed in each specialty on a given day is small. However, including this requirement significantly complicates the solution methodology. Recent theoretical work on integer stochastic dynamic programs include Gade et al. (2014), Kong et al. (2013), and Sun et al. (2015), but there is not much literature on two-stage integer stochastic programming for workforce planning.

Literature related to dynamic optimization based staff planning assume that all the appropriate costs are known. As described before, this is often not the case, since there are several implicit costs in staff planning. Dexter and O'Neill (2001) discuss the importance of implicit costs in creating a staffing plan for anesthesiologists. Therefore, for an optimization model to be useful, these implicit costs have to be estimated and included. In the econometric literature Rust (1987) discusses a structural estimation of the costs involved the dynamic problem of replacement of machinery. Aguirregabiria (1999) describes the approach towards estimation of unknown cost parameters in the joint pricing and inventory management problem at a retail firm. In the operations management literature, Allon et al. (2011) use structural estimation approach to estimate the impact of waiting time performance on the market share in the fast food industry. Deshpande and Arıkan (2012) estimate the impact of airline schedules on flight delays. Structural estimation of operational parameters has also been used in the call center industry by Aksin et al. (2013) and Aksin et al. (2017) to estimate customer preferences. In terms of application context, our paper is closest to Olivares et al. (2008), who model the operating room time allocation problem as a newsvendor problem. They then employ a structural estimation approach to assess the relative costs of idle time and overtime for Operating Rooms (OR). However, all these papers use the estimates created from structural estimation primarily for descriptive purposes and they are not linked with an optimization model. This link is of significant importance in our application context. Furthermore, structural estimation assumes that the decision maker makes optimal decisions and therefore does not capture the errors made by the decision maker in the decision process. To overcome this, in our estimation procedure, we use an approach similar to Su (2008), Ho et al. (2010), and Bolton et al. (2012) who assume that the decision maker is bounded rational. This implies that they are not perfect optimizers and make errors both due to insufficient information and cognitive limitations.

1.3. Contributions

Our paper makes the following contributions. First, we develop a two-stage integer stochastic dynamic programming model for medium and short term planning for anesthesiologists, while incorporating implicit costs, demand uncertainty and service specialties. To the best of our knowledge, this is the first paper to consider this approach in the health care industry. Second, this paper develops a procedure to estimate implicit cost parameters used in the model. This provides a framework for creating staff planning models that overcome the shortcomings of dynamic optimization models in situations where some cost parameters may be implicit, as often the case in service organizations. Third, we provide structural results and develop a general method for solving two-stage integer stochastic dynamic programs. These can also be used in other applications. Fourth, we test our model with real data at the operating services department at the UCLA RRMC, and demonstrate cost savings from such an estimation and optimization approach. We also draw managerial insights from this work.

The remainder of the paper is organized as follows. In Section 2 we provide the formulation of the model and describe the variables, parameters, objectives, and constraints. We also provide structural properties of the model and describe its solution method. In Section 3 we describe the data and methodology for the estimation of demand for anesthesia services from historical data. In Section 4 we present the procedure to estimate the implicit cost parameters. In Section 5, we describe the results from the computational analysis. In Section 6 we summarize our work, provide managerial insights, describe the limitations of our study, and suggest future research directions.

2. Model

We start by presenting a model formulation of the staff planning problem. To provide a precise definition of the model, let S be the set of service specialties $\{$*Cardio-Thoracic, General, Neuro, Pediatric*$\}$, and T be the set of days in a given month. We define the following variables which are optimized.

x_{st}: Number of anesthesiologists of specialty $s \in S$ placed on regular duty on day $t \in T$.

y_{st}: Number of anesthesiologists of specialty $s \in S$ placed on the on-call consideration list on day $t \in T$.

z_{st}: Number of anesthesiologists of specialty $s \in S$ *called* from the on-call list for day $t \in T$.

Next, we define the following parameters or inputs:

n_{st}: The number of anesthesiologists of specialty s available for day $t \in T$.

h: The regular hours of work done per day for an anesthesiologist (hours).

c_o: Overtime cost of anesthesiologists (\$/hour).

c_u: Idle time cost of anesthesiologists (\$/hour).

c_q: Cost of calling an anesthesiologist from the on-call list (\$/day).

c'_q: Cost of keeping an anesthesiologist on the on-call list but not calling. (\$/day).

B_s: The number of hours of anesthesia booked for specialty $s \in S$ for day t. This is a stochastic parameter, which is realized the day before t.

D_{st}: The hours of anesthesia actually realized for specialty $s \in S$. This is a stochastic parameter realized at the end of day t.

$f(D_{st}|B_{st}), F(D_{st}|B_{st})$: the marginal density and distribution of D_{st} given B_{st} respectively.

Further, for conciseness, let:

$a^+ = \max(0, a)$.

$\lceil a \rceil = \min\{n \in \mathbb{Z} | n \geq a\}$.

$\lfloor a \rfloor = \max\{n \in \mathbb{Z} | n \leq a\}$.

$\mathbf{c} = (c_o, c_u, c_q, c_q^{'})$.

The staff planning model is a two-stage, integer stochastic dynamic program. The first stage consists of the Monthly Staff Planning Problem ($MSPP$) which determines the number of anesthesiologist on regular duty and the on-call list for each day of the given month across each specialty. The second stage consists of the Daily Staffing Planning Problem for specialty s in time period t ($DSPP_{st}$). This determines which anesthesiologist to call from the on-call list for specialty s for day t. We next describe each of these problems in detail.

In the $MSPP$, decisions are taken before the beginning of the given month. Thus, at this point the planners are only aware of the historical distribution of B_{st}, and the total number of anesthesiologists available for each day of this month (n_{st}). For each specialty, for each day of the upcoming month, the planners decide the number of anesthesiologists who should be present for regular duty (x_{st}), and the number of anesthesiologists who should be a part of the on-call consideration list (y_{st}). The $MSPP$ is formulated as:

$$(\mathbf{MSPP}) \qquad \mathcal{V}(\mathbf{n},\mathbf{c}) = \min \sum_{s\in S, t\in T} \left\{ \mathbf{E}_{\mathbf{B_{st}}} \left[\mathcal{W}_{st}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st}) \right] \right\} \tag{1}$$

subject to,

$$x_{st} + y_{st} \le n_{st} \qquad \forall s \in S, t \in T \tag{2}$$

$$x_{st}, y_{st} \in \mathbb{N}^+ \qquad \forall s \in S, t \in T \tag{3}$$

The objective (9) represents the total expected monthly costs. This is the sum of expectation of $\mathcal{W}_{st}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ over B_{st}, where the total expectation of the future cost is carried over to the beginning of the horizon when the decision is made. Here, $\mathcal{W}_{st}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ represents the cost of specialty s on day t and depends on the decisions x_{st} and y_{st}, cost parameters $\mathbf{c}$, the number of available anesthesiologists n_{st}, and the booked time B_{st}. The exact form of $\mathcal{W}_{st}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ will be defined in the $DSPP_{st}$. Constraint (2) enforces the total allocation of anesthesiologists for each specialty and each time period cannot be greater than the total availability of anesthesiologists on that day and specialty. Constraint (3) ensures that the decision variables are positive integers.

Next, we describe the second stage problem, $DSPP_{st}$, which considers the daily decision of calling in additional anesthesiologists from the on-call consideration list to support the surgical schedule for next day. At this point the planner is aware of the total booked hours of surgeries for each specialty (B_{st}). Using this information and knowledge of the conditional distribution of the actual realization of surgery duration ($f[D_{st}|B_{st}]$), the planner decides to call in certain number of additional anesthesiologists from the on-call consideration list (z_{st}). Each of these anesthesiologists will be paid an additional amount (c_q). On the day of surgery the actual surgical duration of each surgery is realized, which determines the total workload for each service specialty (D_{st}). Depending on the total available labor hours of each specialty ($h(x_{st} + y_{st})$), the overtime and idle time costs will be realized. The $DSPP_{st}$ is formulated as:

$\mathbf{DSPP}_{st}$

$$\mathcal{W}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st}) = \min \Bigg\{ \left[c_q z_{st} + c'_q (y_{st} - z_{st}) \right] + \mathbf{E}_{D_{st}|B_{st}} \Bigg[c_o \left(D_{st} - h\left(x_{st} + z_{st} \right) \right)^+ +$$

$$c_u\left(h\left(x_{st}+z_{st}\right)-D_{st}\right)^{+}\Bigg]\Bigg\} \tag{4}$$

$$z_{st} \leq y_{st} \tag{5}$$

$$z_{st} \in \mathbb{N}^{+} \tag{6}$$

The objective (4) of the $DSPP_{st}$ consists of four terms. The first term, $c_q z_{st}$, is the cost of extra payments made to the anesthesiologists who are called from the on-call consideration list. The second term, $c_q'(y_{st}-z_{st})$, is the inconvenience cost of not calling $(y_{st}-z_{st})$ anesthesiologists from the on-call consideration list. The third term, $c_o\left(D_{st}-h\left(x_{st}+z_{st}\right)\right)^{+}$ is the overtime payment when the demand realized is greater than the total work load available for specialty s. The fourth term, $c_u\left(h\left(x_{st}+z_{st}\right)-D_{st}\right)^{+}$ is cost of idle time when the demand falls short of total available work hours. For these costs, the expectation is taken over the conditional distribution of D_{st}. Note that the third and fourth term together are the expected costs of the day of surgery and similar to the well known newsvendor cost (Nahmias and Cheng 2009). Constraint (5) restricts the additional number of anesthesiologists who can be called to the ones who are on the on-call consideration list, which is set in the first stage. Constraint (6) restricts the decision variable z_{st} to be a positive integer.

2.1. Structural Properties

In this section we derive structural properties of the model, which can be used to develop its solution method. Let $\mathcal{U}(z_{st})$ denote the objective function of the $DSPP_{st}$, where $\mathcal{U}(z_{st})$ is as given below,

$$\mathcal{U}(z_{st}) = \Bigg\{ \left[c_q z_{st} + c_q'(y_{st}-z_{st})\right] + \mathbf{E}_{D_{st}|B_{st}}\Bigg[c_u\left(D_{st}-h\left(x_{st}+z_{st}\right)\right)^{+} + c_o\left(h\left(x_{st}+z_{st}\right)-D_{st}\right)^{+}\Bigg]\Bigg\} \tag{7}$$

The first proposition provides the optimal solution for the daily staff planning problem $(DSPP_{st})$.

Proposition 1. *If the distribution of $D_{st}|B_{st}$ is stochastically increasing in B_{st}, then there exist thresholds $B^L_{st}(x_{st})$ and $B^U_{st}(x_{st}, y_{st})$ such that the optimal solution for $DSPP_{st}$ is given by $z^*_{st}(x_{st}, y_{st}; B_{st})$:*

$$z^*_{st}(x_{st}, y_{st}; B_{st}) = \begin{cases} \lceil \hat{z}_{st} \rceil & \text{if } \mathcal{U}\left(\lceil \hat{z}_{st} \rceil\right) \leq \mathcal{U}\left(\lfloor \hat{z}_{st} \rfloor\right) \\ \lfloor \hat{z}_{st} \rfloor & \text{otherwise} \end{cases} \tag{8}$$

Where,

$$\hat{z}_{st} = \begin{cases} 0 & \text{if } B_{st} \leq B^L_{st}(x_{st}) \\ \dfrac{1}{h}F^{-1}\left[\dfrac{c_o h + c_q' - c_q}{h(c_u+c_o)}\right] - x_{st} & \text{if } B^L_{st}(x_{st}) \leq B_{st} \leq B^U_{st}(x_{st}, y_{st}) \\ y_{st} & \text{if } B_{st} > B^U_{st}(x_{st}, y_{st}) \end{cases}$$

All proofs are provided in the Appendix. The expressions for threshold values B_{st}^L and B_{st}^U for the lognormal distribution (used to fit the data in the demand estimation procedure in Section 3.2) are described in the proof of Proposition 1. This proposition implies that the number of anesthesiologists who should be called from the on-call can be described as a threshold policy depending on the booked time information B_{st} that is available the day before surgery. If the booked time is below B_{st}^L, then the number of anesthesiologists available on regular duty (x_{st}) would be sufficient. If the booked time is above B_{st}^U, then all the anesthesiologists on the on-call consideration list would be required. For intermediate values of B_{st}, the proposition above provides for the optimal number of anesthesiologists who should be called from the on-call list.

Let $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ be the linear programming relaxation of $\mathbf{DSPP}_{st}$ with the integrality constraint (6) relaxed. Then we define the $\mathbf{MSPP}^{'}$ as:

$$(\mathbf{MSPP}^{'}) \qquad \mathcal{V}^{'}(\mathbf{n}, \mathbf{c}) = \min \sum_{s\in S, t\in T} \left\{ \mathbf{E}_{\mathbf{B_{st}}} \left[\mathcal{W}_{st}^{LP}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st}) \right] \right\} \tag{9}$$

subject to,

$$(2), (3) \tag{10}$$

Since $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st}) \leq \mathcal{W}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$, $\mathcal{V}^{'}(\mathbf{n}, \mathbf{c}) \leq \mathcal{V}(\mathbf{n}, \mathbf{c})$. Thus, the $MSPP^{'}$ is a lower bound to the $MSPP$. The next proposition provides a property of $MSPP^{'}$ that will be used in constructing its solution method.

Proposition 2. *The $MSPP^{'}$ is discrete convex in (x_{st}, y_{st}).*

2.2. Solution Method

Next, we utilize Proposition 1 and 2 to develop a computationally tractable algorithm to solve the $MSPP$. First we solve the integer convex program $MSPP'$. To do so, we approximate the expectation in $MSPP^{'}$ by its sample average approximation (SAA) as:

$$\mathcal{V}^{'}(\mathbf{n}, \mathbf{c}) \approx \hat{\mathcal{V}}'(\mathbf{n}, \mathbf{c}) = \min \frac{1}{M} \sum_{m=1}^{M} \left[\sum_{s\in S, t\in T} \mathcal{W}_{st}^{LP}(x_{st}, y_{st}; \mathbf{c}, b_{st}^m, n_{st}) \right] \tag{11}$$

subject to,

$$(2), (3) \tag{12}$$

As shown in Proposition 2, $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, b_{st}^m, n_{st})$ is a discrete convex function. Therefore, the sample average approximation $\hat{\mathcal{V}}'(\mathbf{n}, \mathbf{c})$ is also a discrete convex problem. We solve $\hat{\mathcal{V}}'(\mathbf{n}, \mathbf{c})$ by solving its integer relaxation employing the sub-gradient method for constrained problems (Boyd et al. 2003). As $\hat{z}_{st} = 0$ is always a feasible solution to $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, b_{st}^m, n_{st})$, there will always be a solution to $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, b_{st}^m, n_{st})$ for every feasible (x_{st}, y_{st}) at each iteration of the sub-gradient method. Furthermore, we stop the sub-gradient method when the current solution does not improve the previous best solution by a pre-specified tolerance. Let this current solution be (x_{st}^*, y_{st}^*) with a corresponding objective value of $(1/M)\sum_{m=1}^{M}\sum_{s\in S,t\in T}\mathcal{W}^{LP}(x^*_{st},y^*_{st};\mathbf{c},b^m_{st},n_{st})$. This value is a lower bound to the $MSPP$. Then, we find the best nearest feasible integer solution $(\hat{x}_{st},\hat{y}_{st})$, and its corresponding objective value $(1/M)\sum_{m=1}^{M}\sum_{s\in S,t\in T}\mathcal{W}(\hat{x}_{st},\hat{y}_{st};\mathbf{c},b^m_{st},n_{st})$. This provides a heuristic solution to the $MSPP$. Define $\hat{g}(\hat{x}_{st},\hat{y}_{st})$, an estimate of the integrality gap at $(\hat{x},\hat{y}_{st})$ as:

$$\hat{g}(\hat{x}_{st},\hat{y}_{st}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{s\in S,t\in T}\mathcal{W}(\hat{x}_{st},\hat{y}_{st};\mathbf{c},b^m_{st},n_{st}) - \frac{1}{M}\sum_{m=1}^{M}\sum_{s\in S,t\in T}\mathcal{W}^{LP}(x^*_{st},y^*_{st};\mathbf{c},b^m_{st},n_{st}) \tag{13}$$

In the above equation, the integrality gap is defined as the difference between the cost of the nearest feasible integer solution from its optimal continuous solution, averaged across M realizations of anesthesia hours by specialty and day. The heuristic algorithm based on SAA to solve the $MSPP$ is formalized below.

Heuristic Algorithm to solve $MSPP$

1. Set $\epsilon > 0$ to be sufficiently small and M to be sufficiently large.

2. For a given (s,t), draw M samples of B_{st}, represented by b^k_{st}, $k = 1,2,\ldots,M$, from the distribution of B_{st}.

3. Solve the convex program $\hat{\mathcal{V}}'(\mathbf{n},\mathbf{c})$ by employing Proposition 2 and the sub-gradient method. Use Proposition 1 to compute $\mathcal{W}^{LP}(x_{st},y_{st};\mathbf{c},b^k_{st},n_{st})$ at each iteration of the sub-gradient method. Let the sub-gradient solution be (x^*_{st},y^*_{st}).

4. Find the best nearest feasible integer solution $(\hat{x}_{st},\hat{y}_{st})$ corresponding to the sub-gradient solution (x^*_{st},y^*_{st}).

5. Compute the estimate of the integrality gap $\hat{g}(\hat{x}_{st},\hat{y}_{st})$ using (13)

6. If $\hat{g}(\hat{x}_{st},\hat{y}_{st}) > \epsilon$, increase sample size M and go to Step 2. Else, $(\hat{x}_{st},\hat{y}_{st})$ is the heuristic solution solution for the given (s,t). Go to Step 7

7. Repeat Step 2 to Step 6 $\forall(s,t)$

8. End.

It is apparent from the above algorithm that the MSPP is decomposable by both specialty and days. Thus, this problem could potentially be solved by more direct methods such as complete enumeration of the first stage variables. However, as discussed in the literature, these complete enumeration methods for two-stage stochastic integer programs could be computationally challenging (Schultz et al. 1998). To test this approach in our context, we decomposed the problems by specialty and found that solving this problem across the four specialties for a given day took about an hour, and for the whole month it took about 98 hours or over four days. This seemed computationally intensive from a practical standpoint. Furthermore, the analysis in Sections 5.2 through

5.5 required solving several instances of the MSPP. Thus, such complete enumeration based methods would preclude these types of analysis, which were important from a practical standpoint. In contrast, our algorithm described above solved the entire problem in less than 10 minutes in all the considered test problem instances, and was within 2% of the costs of the solution obtained by the enumeration based approach. More details are provided in Table EC 13 in the Electronic Companion. Therefore, it seemed reasonable to employ our solution method to solve this problem and conduct the associated analysis.

Finally, it is important to note that the value of the solution using this method would naturally depend on the reliability of the cost parameters, c_q, $c_q^{'}$, c_o, c_u. While c_q and c_o are known, as these are actual dollar payments the hospital makes to the anesthesiologists, c_q' and c_u are implicit. Therefore, we develop an estimation procedure to determine these costs. This procedure first requires estimating the demand distributions for anesthesia services at each specialty. Thus, in the next section, we describe our methodology to specify and estimate these distributions.

3. Estimation of Demand Distributions

Estimation of demand distribution for anesthesia services consists of two stages. First we estimate the distribution for the booked hours for specialty s and day t (B_{st}). We then estimate $D_{st}|B_{st}$, the distribution for the daily anesthesia hours used for specialty s on day t, conditional on the booked hours B_{st}.

3.1. Estimating Distribution of Booked Hours (B_{st})

Surgery requests start coming in sequentially about six months before the day of surgery. Subsequently, there are requests for cancellations and add-on cases that keep coming in until one day before the day of surgery. While these advance bookings might be informative about the actual realization of B_{st}, this information is not passed on by the other hospital departments to the operating services department as they are subject to change. Only the final booked hours for each department is sent by admissions to operating services the day before the scheduled surgeries. This implies that, no advanced information from early bookings is available when the $MSPP$ is being solved. The information available is restricted to the day of week, month, and whether upcoming day is a holiday. Therefore, we use only these variables to estimate the distribution of B_{st}. In Figure 1 we plot the empirical distribution of the booked hours for each of the specialties (B_{st}).

Insert Figure 1 here

From Figure 1 we can see that for Cardio-Thoracic, Neuro, and Pediatric surgeries there is a concentration of data at zero. This is because these surgeries are more specialized and they are not performed every day of the week. General surgeries on the other hand are performed almost every day and we do not see such concentration of data at zero. Therefore, we used separate estimation procedure for specialized and general surgeries. We next describe these methods.

Estimation of B_{st} for specialized surgeries

In order to estimate the distribution of booked anesthesia hours for specialized surgeries such as Cardio-Thoracic, Neuro, and Pediatric surgeries we use a two-step estimation method. A more detailed description of this method can be found in Duan et al. (1983) and Min and Agresti (2002). Here, in the first step, the dependent variable is a binary outcome variable with $B_{st}=0$ indicating there is no demand for specialty s on day t. Conditional on this first stage binary variable being false (i.e. $B_{st}>0$), we then estimate the magnitude of B_{st}.

More specifically, in the first step, the binary outcome variable B_{st} is modeled by logistic regression. The specification of this logistic regression is:

$$\text{logit}[P(B_{st})=0]=\alpha_{s,0}+\alpha_{s,1}\times Day\ of\ Week_t+\alpha_{s2}\times Month_t+\alpha_{s3}\times Holiday_t \tag{14}$$

This can be written concisely as:

$$\text{logit}[P(B_{st})=0]=\boldsymbol{\alpha}_s^{'}\mathbf{h}_t\,. \tag{15}$$

In the second part of the estimation procedure, we estimate the distribution of the magnitude of B_{st}, conditional on it being positive. We use a lognormal specification of the magnitude of B_{st} for better fit. A lognormal distribution for surgical services demand has been used by Duan et al. (1983), May et al. (2000), and He et al. (2012). This specification is:

$$\log(B_{st}|B_{st}>0)=\beta_{s0}\times Day\ of\ Week+\beta_{s1}\times Month+\beta_{s2}\times Holiday+\epsilon_{st} \tag{16}$$

We simplify the above as,

$$\log(B_{st}|B_{st}>0)=\boldsymbol{\beta}_s^{'}\mathbf{h}_t+\epsilon_{st}, \tag{17}$$

where $\epsilon_{st}\sim\mathcal{N}(0,\sigma_s^2)$. Following Duan et al. (1983) and Min and Agresti (2002), the maximum likelihood of two part model is given by,

$$\ell(\boldsymbol{\alpha_S},\boldsymbol{\beta_S},\sigma)=\ell_1(\boldsymbol{\alpha_S})\ell(\boldsymbol{\beta_S},\sigma) \tag{18}$$

Where,

$$\ell_1(\boldsymbol{\alpha_S})=\left[\prod_{B_{st}=0}e^{\boldsymbol{\alpha}_s^{'}\mathbf{h}_t}\right]\left[\prod_{t=1}^{n}\frac{1}{1+e^{\boldsymbol{\alpha}_s^{'}\mathbf{h}_t}}\right] \tag{19}$$

and

$$\ell_2(\boldsymbol{\beta}_s,\sigma_s)=\prod_{B_{st}>0}\sigma_s^{-1}\phi\left(\frac{\log(B_{st})-\boldsymbol{\beta}_s^{\prime}\mathbf{h}_t}{\sigma_s}\right). \tag{20}$$

As the likelihood function is separable in the parameters, we can estimate $\boldsymbol{\alpha}_s,\boldsymbol{\beta}_s$, and σ by independently solving the maximum of the two likelihood functions, $\ell_1(\boldsymbol{\alpha}_s)$ and $\ell_2(\boldsymbol{\beta}_s,\sigma_s)$.

We summarize the results of the estimation procedure in the Electronic Companion. From these results we can conclude that the procedure is very effective in estimating B_{st} for specialized surgeries at the UCLA RRMC.

Estimation of $B_{s,t}$ for General Surgeries

We can observe from Figure 1 that the distribution of booked anesthesia hours for general surgeries is bimodal. This is because, while general surgeries are performed on most days, there is lower demand on weekends and holidays, while there is higher demand on regular days. Therefore, we model the distribution of anesthesia booked for general surgeries as a mixture of two Gaussian distributions. This approach for modeling bimodal distributions has been suggested by Allenby et al. (1998) for capturing a wide variety of heterogeneity in demand distributions. In Gaussian mixture models, the distribution of the mixture is given by the weighted sum of the two Gaussian distributions. Thus, the conditional distribution $g(B_{st}|\mathbf{h}_t)$ is given by

$$g(B_{st}|\mathbf{h}_t) = \sum_{k\in\{1,2\}} \pi_k \phi_k(B_{st}|\mathbf{h}_{tk};\boldsymbol{\beta}_k), \tag{21}$$

where, π_k are weights assigned to the two component distributions and $\phi_k(B_{st}|\mathbf{h}_{tk};\boldsymbol{\beta}_k)$ are the two component distributions with regression parameters $\mathbf{h}_{t1}$ and $\mathbf{h}_{t2}$, and coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. We estimate this Gaussian mixture model using the `flexmix` package in R (Grün and Leisch 2007). The results of the two component regressions are summarized in the Electronic Companion. Here again, these results show that this is an effective procedure to estimate B_{st} for general surgeries at the UCLA RRMC.

3.2. Estimation of $D_{st}|B_{st}$

We choose a lognormal specification for $F(D_{st}|B_{st})$ as it provides a good fit (as shown in the Electronic Companion). In addition, the lognormal specification has been used in the literature for modeling of demand for surgical services (Strum et al. 2000, He et al. 2012). The specification of the regression model for D_{st} is given as:

$$\log(D_{st}) = \gamma \log(B_{st}) + \xi_s \quad \forall s \in S, t \in T. \tag{22}$$

Here, $\xi_s \sim \mathcal{N}(0, \sigma_s^{'2})$. We present the results of the estimation of $D_{st}|B_{st}$ across each specialty in the Electronic Companion. These results validate the choice of lognormal specification to estimate $D_{st}|B_{st}$.

4. Estimation Procedure for Implicit Cost Parameters

To estimate the implicit cost parameters we adapt the approach followed in the estimation of discrete choice models (McFadden 1974, McFadden and Manski 1981). To enable this, we assume that the staff planner does not know the numerical value of the implicit costs, but is aware of the cost trade-offs when making staff planning decisions. Therefore, the planner has subconscious relative weights in mind and uses these costs imperfectly. We observe the historical daily decisions of the staff planner on how many anesthesiologists were actually called from the on-call consideration list. We then employ a maximum likelihood optimization to estimate the implicit cost parameters in a manner that best explains the staff planner's decisions observed in the data. The estimation procedure for implicit cost parameters consist of the following steps:

1. We develop a decision model of the staff planner.
2. Based on this decision model, we derive the likelihood of obtaining the observed data as a function of the cost parameters.
3. Finally, we choose the implicit cost parameters which maximizes the likelihood of observing the data.

We next describe each step in detail.

4.1. Decision Problem of Staff Planner

The literature related to operating room staff planning shows experimental evidence that operating room planners demonstrate errors and biases from the optimal solution (Wachtel and Dexter 2010). Therefore, we model the staff planner as a bounded rational decision maker, who is not a perfect optimizer, but makes errors due to the limited availability of information or because of cognitive limitations. Furthermore, consistent with quantal choice theory (McFadden 1976), we assume that when the planner is faced with alternative staff planning options, instead of the selecting the optimal staffing plan, they select better options with higher probability.

The above evidence that the staff planner is a bounded rational decision maker precludes the use of data on the monthly decisions for estimating the cost parameters. The monthly decision of the staff planner, on deciding which of the available anesthesiologists should be placed on regular duty and who should be on the on-call consideration list, is a two-period stochastic dynamic problem. Thus, modeling the monthly decisions of the staff planner would require a structural model of *dynamic* discrete choices. Estimating parameters in dynamic discrete choices requires the assumption that the decision maker is a rational agent. In the literature related to structural estimation of dynamic discrete choices this is a standard assumption and referred to as the rational expectations assumption (Aguirregabiria and Mira 2010). Since we assume that the staff planner is not rational, but is bounded rational and makes errors in their staff planning, we do not assume rational expectations and exclude the monthly data in our estimation procedure.

Alternatively, we use data on daily decisions and the logit choice model to evaluate the probability with which the staff planner selects the number of anesthesiologists to call from the on-call consideration list. The logit model suitable in our context for two reasons. First, it allows for the discrete choices, like number of anesthesiologists. Second it leads to an analytically tractable maximum likelihood model. Our context is similar to Su (2008), who uses the multinomial logit choice model and provides empirical evidence that a logit choice model provides a good fit for a bounded rational newsvendor.

According to logit choice model, the probability of selecting a decision x is proportional to $e^{U(x)}$, where $U(x)$ is the utility of selecting the decision x (McFadden 1974). Consequently, if the domain of decisions is X, the probability of selecting choice x is given by:

$$p(x) = \frac{e^{U(x)}}{\sum_{x \in X} e^{U(x)}} \quad . \tag{23}$$

Next, we use the above logit choice probability to derive the likelihood of the staff planner calling a certain number of anesthesiologists from the on-call consideration list.

4.2. Deriving the Likelihood Function for Staff Planning Decisions

For conciseness, we represent $\mathcal{U}(\mathbf{c}, z_{st}, y_{st} B_{st})$ as follows:

$$\mathcal{U}(\mathbf{c}, z_{st}, y_{st}, B_{st}) = \Bigg\{ \left[c_q z_{st} + c_q'(y_{st} - z_{st})\right] + \mathbf{E}_{D_{st}|B_{st}} \Bigg[c_u \left(D_{st} - h\left(x_{st} + z_{st}\right)\right)^+ + c_o \left(h\left(x_{st} + z_{st}\right) - D_{st}\right)^+ \Bigg] \Bigg\} \tag{24}$$

For the daily staff planning the utility of calling z_{st} anesthesiologists from the on-call consideration list, for a given choice of cost parameter $\mathbf{c}$, booked time B_{st} and y_{st} over all other feasible $z_{st}^{'}$, is given as the negative of the cost incurred, or, $-\mathcal{U}(\mathbf{c}, z_{st}, y_{st} B_{st})$. Therefore, from (23), the probability of choice z_{st} is:

$$p_{st}(\mathbf{c}, z_{st}, y_{st} B_{st}) = \frac{\exp(-\mathcal{U}(\mathbf{c}, z_{st}, y_{st} B_{st}))}{\sum_{z_{st}' \leq y_{st}} \exp(-\mathcal{U}(\mathbf{c}, z_{st}', y_{st} B_{st}))} \tag{25}$$

Therefore, the likelihood of observing z_{st} for all s, t in the data for a given choice of $\mathbf{c}$ will be given by:

$$\mathcal{L}(\mathbf{c}) = \Pi_{s \in S} \Pi_{t \in T} p_{st}(\mathbf{c}, z_{st}, y_{st} B_{st}) \tag{26}$$

4.3. Determining Costs to Maximize the Likelihood Function

Maximizing the likelihood function as described in (26), is challenging because computing the likelihood requires multiplication of $|S| \times |T|$ probabilities. The resultant likelihood becomes extremely small and we run into floating point errors when this function is maximized. In order to mitigate this, it is common practice to maximize the log-likelihood (Cameron and Trivedi 2005). Since the logarithm function is monotonically increasing, the optimal solution will not change. The estimate of $\mathbf{c}$ which maximizes the log-likelihood is given by:

$$\hat{\boldsymbol{c}} = \arg\max_{\mathbf{c}} \log \mathcal{L}(\mathbf{c}). \tag{27}$$

Using (26), this simplifies to:

$$\hat{\boldsymbol{c}} = \arg\max_{\mathbf{c}} \sum_{s \in S, t \in T} \log \left\{ p_t(\mathbf{c}, z_{st}, y_{st} B_{st}) \right\}. \tag{28}$$

We first show that the above optimization problem is concave in $\mathbf{c}$ and then propose an estimation procedure.

Proposition 3. $\log \mathcal{L}(\mathbf{c})$ *is concave in* $\mathbf{c}$.

In light of Proposition 3, a local solution of a non-linear solver would be the global optimum. We use the non-linear solver NLOPT (Johnson 2014) with a Python programming interface to solve the maximum likelihood problem for a given dataset. Additionally, for computational stability, during the non-linear optimization, we normalize c_q to 1. We also employ a non-parametric bootstrap analysis for our estimation procedure. Bootstrap analysis allows us to compute an approximation of the confidence interval of the cost estimates. To perform bootstrap analysis, we follow the procedure described in (Greene 2003). We take J samples with replacement from our data set. For each sample we compute the cost estimates by solving equation (28) for the sampled dataset. Thus, we have J cost estimates $\{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_J\}$. The mean of the cost estimates is given by, $\bar{\mathbf{c}} = \frac{1}{J}\sum_j \hat{\mathbf{c}}_J$, and we use the 2.5^{th} and 97.5^{th} percentile of these cost estimates to obtain the 95% confidence interval of the estimates. We report these values in Table 2. Note that the cost estimates are scaled such that $c_q = 1$.

Insert Table 2 here

We observe in Table 2 that the estimated cost of not calling an anesthesiologist on the on-call consideration list, is 1.56 times the cost of actually calling the anesthesiologists. This seems plausible as the anesthesiologist loses not only the additional income from call, but potentially foregoes the opportunity to make income from other sources during that day. Dexter and O'Neill (2001) discuss the impact of these implicit costs on on-call staffing, but such costs have not been quantified in the literature thus far. Incorporating such costs are important, as otherwise, this would lead to a long on-call consideration list. While this may provide the staff planner the flexibility to react to updated information without incurring supplemental financial expenses at the hospital, this would lead to larger schedule unpredictability for the anesthesiologists. Such schedule unpredictability has been associated with higher employee dissatisfaction in healthcare (Gander et al. 2007), and in other industries like retail (Henly and Lambert 2014). Higher dissatisfaction can lead to increased employee turnover, which could be detrimental to the hospital.

Additionally, when we scale c_q to 1, the corresponding value of the explicit costs of overtime $c_o = 0.18$. This implies, that the idle cost of an anesthesiologist is 1.94 times the overtime cost. This result is consistent with Olivares et al. (2008) who find that the cost of OR idle time was observed to be 60% higher than the cost of OR overtime. Our study demonstrates that a similar effect is in place for managing on-calls for anesthesiologists.

To better understand how the estimates of implicit costs changed with factors such as the data time frame, day of the week and specialty, we conducted additional analysis, which are summarized in the Electronic Companion. From this analysis, we can conclude that the implicit costs were quite stable and did not vary significantly with these factors. This shows that the staff planning decisions were made in a consistent manner at the operating services department at this hospital, and no specialty was preferred over the other. This is very desirable from the perspective of staff morale.

Finally, the staff planner's problem can be broadly considered as a newsvendor problem with overstock costs corresponding to the implicit costs of not calling an anesthesiologist from the on call list and the costs of idle capacity. Similarly, the understock costs will be the costs of calling an anesthesiologist from the call list and the costs of overtime. Whenever, there are such newsvendor trade-offs between overstock and understock costs, decision makers have shown to exhibit systematic biases (Schweitzer and Cachon 2000, Bostian et al. 2008, Ho et al. 2010). One such common and well-studied bias is anchoring decisions on mean demand. This means that instead of ordering the optimal expected profit maximizing quantity, decision makers order a quantity between the optimal quantity and the quantity required to meet the mean demand (Bostian et al. 2008). Wachtel and Dexter (2010) also discuss the possibility of staff planners for anesthesiologists demonstrating anchoring on mean demand. As described in the Electronic Companion, using the approach in Bostian et al. (2008), we also found evidence to indicate that the staff planners decisions could be driven by a mean anchoring bias. Quantile choice theory has been used to explain the mean anchoring bias in several applications in operations management (Chen and Song 2019). Thus, this provides more validation to represent the staff planners decisions using this theory.

5. Computational Analysis

In this section, we first perform computational analysis to validate the performance of the estimation procedure described in Section 4. Then, we show the benefits of using the solution method described in Section 2.2 over current practice. We also use our model to evaluate the impact of changes in costs, booked time variability and the impact of hiring more anesthesiologists for particular specialties.

5.1. Validation of estimated cost parameters

In order to validate the cost estimation procedure, we demonstrate that our model can accurately predict the decisions of the staff planner using the estimated costs. We follow a 10-fold cross validation procedure to quantify the prediction accuracy of our model. Kohavi et al. (1995) provide a detailed discussion on the advantages of using of k-fold models for cross validation. They propose $k = 10$ for discrete models such as the multinomial logit. In a 10-fold cross validation approach we divide our data set Δ into ten mutually exclusive subsets (folds) $\{\Delta_1, \dots, \Delta_{10}\}$ of approximately equal size. We then use the estimation procedure (described in Section 4) ten times. Each time the cost parameters are estimated using dataset $\Delta \setminus \Delta_i$. Let these estimated parameters be $\hat{\boldsymbol{c}}_i$.

Next, given these estimates we use equation (25) to compute the predicted choice probability $\hat{p}_{st}(\hat{\boldsymbol{c}}_i, z_{st}, y_{st}, B_{st})$ for each feasible z_{st} for the data set Δ_i. Then, because the staff planner's choice is modeled as a multinomial logit, the predicted decision of the staff planner will be the decision which has the highest predicted probability. Thus, the predicted decisions for the test data set Δ_i will be:

$$\hat{z}^i_{st} = \arg\max_{z_{st}} \{\hat{p}_{st}(\hat{\boldsymbol{c}}_i, z_{st}, y_{st}, B_{st})\} \quad \forall (s,t) \in \boldsymbol{\Delta}_i \forall i \in \{1, 2, \ldots, 10\}. \tag{29}$$

We compute the Root Mean Square Error ($RMSE$) of the above predicted decisions $\hat{z}^i_{st}$ with respect to the actual historical decisions of the staff planner $\tilde{z}^i_{st}$ for each of the 10 datasets Δ_i. Then, we compute the average RMSE across the 10 sets of predictions as:

$$\overline{RMSE} = \frac{1}{10} \sum_{i=1}^{10} \sqrt{\frac{\sum_{(s,t)\in\boldsymbol{\Delta}_i} (\tilde{z}^i_{st} - \hat{z}^i_{st})^2}{|\boldsymbol{\Delta}_i|}} \ . \tag{30}$$

We also compute the accuracy of the model as the percentage of times the model predicted the correct decision. If $\hat{z}^i_{st} = \tilde{z}^i_{st}$, we denote $I_{\hat{z}^i_{st} = \tilde{z}^i_{st}} = 1$. Therefore the accuracy for the dataset Δ_i is $acc_i = \frac{1}{|\Delta_i|} \sum_{s,t\in\Delta_i} I_{\hat{z}^i_{st} = \tilde{z}^i_{st}}$. The average of the accuracy across the 10 folds would be $\overline{acc} = \frac{1}{10} \sum_{i=1}^{10} acc_i$. We found that the estimation procedure is able to exactly predict z_{st} about 47% of the time. In addition, the error in prediction accuracy was also small, with the average RMSE around 0.67. We also modeled the staff planners decision to determine how many anesthesiologists called from the on-call list (z_{st}) as a linear regression of the observable characteristics like the number of anesthesiologists on regular duty (x_{st}), the number of anesthesiologists on the on-call list (y_{st}), and the total booked hours for surgery (B_{st}). Estimating operational parameters assuming a linear managerial decision rule has been applied previously in Foreman et al. (2010). The results, summarized in the Electronic Companion, showed that the average RMSE for the linear fit is 0.89. This shows that the logit choice model is a better fit to model the staff planners decisions as it better captures the non-linear dependence of z_{st} on y_{st}, x_{st} and B_{st}. This in turn provides validity for the implicit cost estimation procedure described in Section 4.

5.2. Comparison of decisions and costs with current practice

We use the estimated implicit costs to fully specify the $MSPP$ and $DSPP_{st}$. We can now compute the total costs of using a model based solution and compare this to the cost incurred by current practice. When calculating the cost benefits of using the model based solution described in Section 2.2 with respect to the actual decisions of the staff planner, we first define the *ex-post* cost of a decision (x_{st}, y_{st}, z_{st}) as:

$$\mathfrak{U}(x_{st}, y_{st}, z_{st}) = \left\{ \left[c_q z_{st} + c'_q(y_{st} - z_{st})\right] + \left[c_u \left(\widetilde{D}_{st} - h\left(x_{st} + z_{st}\right)\right)^+ + c_o \left(h\left(x_{st} + z_{st}\right) - \widetilde{D}_{st}\right)^+\right]\right\} \tag{31}$$

Here, $\mathfrak{U}(x_{st}, y_{st}, z_{st})$ is the cost when decisions (x_{st}, y_{st}, z_{st}) are taken for day t, and the actual realization of the total durations of surgeries of specialty s is $\widetilde{D}_{st}$.

Let, $(x^m_{st}, y^m_{st}, z^m_{st})$ be the decisions computed by the model based solution procedure described in Section 2.2, and $(\tilde{x}_{st}, \tilde{y}_{st}, \tilde{z}_{st})$ are the actual decisions of the staff planner. We employ $\mathfrak{U}(x_{st}, y_{st}, z_{st})$ to compare the benefits of the model based solutions to the actual decisions of the staff planner by calculating the percentage relative cost improvement as:

$$\delta_{st} = 100\% \times \frac{|\mathfrak{U}(x^m_{st}, y^m_{st}, z^m_{st}) - \mathfrak{U}(\tilde{x}_{st}, \tilde{y}_{st}, \tilde{z}_{st})|}{\mathfrak{U}(\tilde{x}_{st}, \tilde{y}_{st}, \tilde{z}_{st})} \tag{32}$$

We report the average cost improvement by specialty and overall average cost improvement in Table 3. It can be observed from this table that the average cost saving, observed on historical data, from using the model based solution is 13.72%. Additionally, we observe that the cost improvement from using the model based solution is observed across all the specialties.

Insert Table 3 here

To better understand the reason for this improvement we compared the model based solution with the staff planner's plan in more detail. The results summarized in Table 4 show that on average, the model based solution assigns more anesthesiologists to the on-call consideration list. While this allows for greater flexibility to react to the uncertainty in the booked time (B_{st}), there are costs to having more flexibility. However, the model based solution manages to still reduce overall costs because it creates an on-call consideration list for fewer days that the staff planner, as shown in Table 4.

Insert Table 4 here

Finally, for this model to be accepted by the anesthesiologists, it is important it captures the implicit costs considered by the staff planner and these costs have to be consistent with past practice. Our maximum likelihood procedure estimates these costs from the past decisions of the staff planner. This provides reassurance to the anesthesiologists that we have not only captured implicit costs, but estimated its value based on past decisions which were acceptable to them. In addition, the optimization approach more precisely balances the implicit and explicit costs, which leads to lower total costs. As noted above, compared to past practice, our model reduces total costs by 13.72%. Here, the reduction in the implicit cost component was 33.39%, which was greater than the % decrease in total costs. Thus, the model based solution should be at least as acceptable as the solution provided by the staff planner.

5.3. Impact of changes in cost

Anesthesiologists are one of the most expensive labor categories in the United States, and the mean annual wage has undergone an increase of 14% between 2016-2017 (Bureau of Labor Statistics 2018). Increases in salaries imply a proportional increase in on-call and overtime payments. Our model based solution allows us to evaluate the impact of these cost increases. In Figure 2 we plot the impact of the change in on-call and overtime costs. From this figure, as expected, we can see that the total cost increases with the on-call and overtime costs. However, we can also observe that on a percentage basis, the overall cost is more sensitive to changes in the overtime cost than the on-call cost. This is because overtime costs are incurred on more days as compared to on-call costs. Thus, a percentage changes in overtime cost leads to a higher relative change in the overall cost.

Insert Figure 2 and 3 here

5.4. Impact of changes in booked time variability

In this section we analyze the change in cost when the variability of booked time (B_{st}) is reduced. To isolate the impact of reducing the variability of B_{st}, we take the statistical model of B_{st} as described in equation (16). Then we systematically reduce the standard deviation σ_s, and create simulations of B_{st}. We compute the model based optimal solutions and the *ex-post* cost based on simulations of this modified model of B_{st}.

As shown in Figure 3, even a relatively modest reduction in the standard deviation of B_{st} can lead to cost reductions. In practice, reduction in the variability B_{st} would just require additional information to better forecast demand a month in advance. Some of the ways the UCLA RRMC can potentially do this include incorporating early booking information when deciding the monthly staff planning (Tiwari et al. 2014) and using pre-operative consultations, text, and phone reminders to reduce no-shows (Knox et al. 2009, Milne et al. 2006, Haufler and Harrington 2011). All these initiatives could potentially reduce variability in booked durations without significant capital investment and still reduce overall costs. Our model provides an impetus for doing this by quantifying the benefits of booked time variability reduction.

5.5. Impact of hiring anesthesiologists by specialty

We can use the $MSPP$ to evaluate the impact of hiring additional anesthesiologists across specialties. For this analysis, we systematically increase n_{st} for each specialty s and compute the cost as defined in (31). The results for each specialty are shown in Figure 4. We can observe from the figure that across all specializations overall costs are reduced but there are decreasing returns to scale from hiring additional anesthesiologists. This is because at some point, there is not enough hours of anesthesia required and additional anesthesiologists do not provide any benefit. We can also see from Figure 4 that the marginal benefit of hiring an additional anesthesiologist is highest for general surgeries, followed by Cardio-Thoracic, Neuro and Pediatric surgeries. Therefore, this analysis can help the hospital management prioritize hiring decisions by specialty.

Insert Figure 4 here

6. Conclusions

In this paper we consider the anesthesiologist staffing problem typically found in large multi-specialty hospitals. In this problem, the planner makes monthly and daily staffing decisions on the number of anesthesiologists across each specialty to minimize overall costs. We model the staff planning problem as a two-stage integer stochastic dynamic program, provide its structural properties and use this to develop a sample average approximation based algorithm to solve this model.

While some of the cost components of this model are explicitly known, other cost components are implicit. We assume that the staff planner is aware of the trade-offs between explicit and implicit costs, but is not a perfect optimizer and makes errors in their decisions. To capture this, we develop a decision model of a bounded rational staff planner. Using this decision model, and available historical data of decisions taken by the staff planner, we estimate the implicit costs. This leads to a fully specified model of staff planning. We then compare the costs of the model based solution with the costs resulting from the historical decisions of the staff planner. Based on this analysis, we find that our approach can potentially save around 13% in costs, which translates to about $1.8 million on an annual basis. Our model outperforms current practice as it typically creates an on-call list with more anesthesiologists, thus offering additional flexibility to reduce overtime costs. However, it creates this list only when overall costs are reduced. This leads to our model creating an on-call list on fewer days than the staff planner.

In addition, the estimated costs and the optimization model has generated several managerial insights. First, we observe that the cost of not calling an anesthesiologist on the on-call list is considered 56% more expensive than actually calling the anesthesiologist. Similarly, one hour of idle time was considered 94% more expensive than one hour of overtime. This showed that idling costs are considered to be more expensive than overtime costs and this is consistent with operating room staffing literature (Olivares et al. 2008). Second, our model permitted us to evaluate the impact of reducing the variability of the booked time. We saw that reducing the variability can potentially lower costs by up to 12% by better predicting the requirements for anesthesia services. We were also able to evaluate the overall effect of changing overtime and on-call costs and the benefits of hiring additional anesthesiologists by specialty.

Our study has the following limitations. First, it is possible that there are some unobserved heterogeneity across individual anesthesiologists, depending on seniority, or other factors. Some anesthesiologists may have a higher cost of not getting called or have costlier idle time. While, it is possible to incorporate this heterogeneity and estimate the different costs across the individual anesthesiologists, we were restricted by our lack of availability of data at the individual anesthesiologist level. Second, in the current staffing plan, the monthly plan is adjusted only once and this is done the day before the surgery. However, it may be possible to update the staff planning when each elective procedure is booked. This has been suggested by Tiwari et al. (2014) and Xie and Zenios (2015). In such a dynamic schedule updating framework there will also be implicit costs. Our procedure can potentially be extended to evaluate these implicit costs. However, we were unable to perform this analysis because the UCLA RRMC only recorded the booking data when it was finalized, the day before the procedures. Finally, our analysis on the impact of hiring anesthesiologists by specialty is restricted to the costs considered in the model. However, there could be additional costs of hiring anesthesiologists such as recruitment costs, bonuses and on-boarding costs. Furthermore, the decision to hire anesthesiologists of certain specializations would also depend on the longer term strategic focus of the hospital towards attracting demand for certain kinds of procedures, or hiring faculty physicians of certain specializations for meeting teaching requirements at the medical school. Since we did not have information on these aspects and the additional costs, we were unable to conduct a more comprehensive and longer term analysis to determine the right sizing of the anesthesiology staff by specialty.

This paper opens up several opportunities for future research. First, we could extend this framework to other industries outside of healthcare. While this paper adds to the evidence that idle time is considered more expensive in the healthcare context, it is not obvious if that is true for other industries like retail, call centers, and airlines that have overtime, on-call and idle time costs. Second, as described above, we can extend our framework to the the context of dynamic staff planning, where staff planning has more than two stages. However, this will require significant modifications to the model and solution procedure.

Appendix

Proofs of Propositions

Proof of Proposition 1

We begin by proving that the integer relaxation of $DSPP_{st}$ is convex in z_{zt}. The objective function of $DSPP_{st}$ is given by:

$$\mathcal{U}(z_{st}) = \left[c_q z_{st} + c_q'(y_{st} - z_{st})\right] + \mathbf{E}_{D_{st}|B_{st}}\Bigg[c_u\left(D_{st} - h\left(x_{st} + z_{st}\right)\right)^+ + c_o\left(h\left(x_{st} + z_{st}\right) - D_{st}\right)^+\Bigg] \tag{33}$$

The first term $\left[c_q z_{st} + c_q'(y_{st} - z_{st})\right]$ is linear in z_{st}. The second term is the newsvendor cost function, which is convex in z_{st} (Nahmias and Cheng 2009). Therefore, the integer relaxation of $DSPP_{st}$ is convex in z_{st}.

From the first order condition,

$$\frac{d\mathcal{U}(z_{st})}{dz_{st}} = 0 \tag{34}$$

Substituting $\mathcal{U}(z_{st})$ and writing the expectation as integration, the above simplifies to,

$$
\begin{aligned}
c_q - c_q^{'} + \frac{d}{dz_{st}}\Bigg\{ & \int_{h(z_{st}+x_{st})}^{\infty} c_o\left[D_{st} - h(x_{st}+z_{st})\right] f(D_{st}|B_{st})dD_{st} \\
& + \int_{0}^{h(z_{st}+x_{st})} c_u\left[h(x_{st}+z_{st}) - D_{st}\right] f(D_{st}|B_{st})dD_{st}\Bigg\} = 0
\end{aligned} \tag{35}
$$

Differentiating under the integral sign and solving for z_{st}, we get,

$$
z_{s,t}^{*} = \frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} \tag{36}
$$

Where $F_{D_{st}|B_{st}}$ is the CDF of $D_{st}|B_{st}$.

As z_{st} is constrained to be positive and less than y_{st}, therefore the following conditions hold:

$$
\frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} \leq 0 \implies z_{st}^{*} = 0 \tag{37}
$$

$$
\frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} \geq y_{st} \implies z_{st}^{*} = y_{st} \tag{38}
$$

$$
0 \leq \frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} \leq y_{st} \implies z_{st}^{*} = \frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} \tag{39}
$$

Expressing $\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}$ as $\kappa(\mathbf{c})$. From (37) we have $z_{st} = 0$ if,

$$
F_{D_{st}|B_{st}}(hx_{st}) \geq \kappa(\mathbf{c}) \tag{40}
$$

Assuming $D_{st}|B_{st}$ is stochastically increasing in B_{st}, then, there exists, some $B_{st}^{L}(x_{st})$ such that for $B_{st} \leq B_{st}^{L}$, $F_{D_{st}|B_{st}}(hx_{st}) \geq \kappa(\mathbf{c})$, and therefore, $z_{st}^{*} = 0$. Similarly, from (38), there exists some $B_{st}^{H}(x_{st}, y_{st})$, such that if $B_{st} \geq B_{st}^{H}(x_{st}, y_{st})$, $F_{D_{st}|B_{st}}(h(x_{st} + y_{st}) \leq \kappa(\mathbf{c})$, and therefore, $z_{st}^{*} = y$. Thus, the solution to the integer relaxation of $DSPP_{st}$ can be written as:

$$
\hat{z}_{st} = \begin{cases} 0 & \text{if } B_{st} \leq B_{st}^{L}(x_{st}) \\ \frac{1}{h} F_{D_{st}|B_{st}}^{-1}\left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)}\right] - x_{st} & \text{if } B_{st}^{L}(x_{st}) \leq B_{st} \leq B_{st}^{U}(x_{st}, y_{st}) \\ y_{st} & \text{if } B_{st} > B_{st}^{U}(x_{st}, y_{st}) \end{cases}
$$

Since the integer relaxation of the $DSPP_{st}$ is convex in z_{st} its integer solution will be:

$$
z_{st}^{*}(x_{st}, y_{st}; B_{st}) = \begin{cases} \lceil \hat{z}_{s,t} \rceil & \text{if } \mathcal{U}\left(\lceil \hat{z}_{st} \rceil\right) \leq \mathcal{U}\left(\lfloor \hat{z}_{st} \rfloor\right) \\ \lfloor \hat{z}_{st} \rfloor & \text{otherwise} \end{cases} \tag{41}
$$

As discussed in Section 3.2, we model $D_{st}|B_{st}$ as a lognormal distribution. Therefore, we derive closed form expressions for $B_{st}^{L}(x_{st})$ and $B_{st}^{U}(x_{st})$ for the case when $D_{st}|B_{st}$ has a lognormal distribution. Here, $F(x_{st}) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{\ln(x_{st}) - \mu}{\sqrt{2}\sigma}\right)$, where μ is the mean of the log of the random variable. In this case $\mu = \mathbf{E}[\ln D_{st}|B_{st}]$. Therefore, from (37), $z_{st} = 0$ if,

$$
\frac{1}{2} + \frac{1}{2} erf\left(\frac{\ln(hx_{st}) - \mathbf{E}[\ln D_{s,t}|B_{s,t}]}{\sqrt{2}\sigma}\right) \geq \kappa(\mathbf{c}) \tag{42}
$$

Simplifying,

$$\ln(hx_{st}) - \mathbf{E}[\ln D_{st}|B_{st}] \geq \sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right) \tag{43}$$

Replacing $\mathbf{E}[\ln D_{st}|B_{st}]$ with $\gamma_{st} \ln B_{st}$ and simplifying,

$$B_{s,t} \leq \left(\frac{hx_{st}}{\sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right)} \right)^{1/\gamma_{s,t}} \tag{44}$$

Therefore for feasibility of $DSPP_{st}$,

$$B_{s,t} \leq \left(\frac{hx_{st}}{\sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right)} \right)^{1/\gamma_{s,t}} \implies z_{st} = 0 \tag{45}$$

Therefore,

$$B^L_{st}(x_{st}) = \left(\frac{hx_{st}}{\sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right)} \right)^{1/\gamma_{s,t}} \tag{46}$$

The second constraint on z_{st} is $z_{st} \leq y_{st}$. Therefore, from (38) for feasibility,

$$\frac{1}{h} F^{-1} \left[\frac{c_o h + c_q^{'} - c_q}{h(c_u + c_o)} \right] - x_{st} \leq y_{st} \tag{47}$$

Employing a similar approach, this constraint can be simplified as:

$$B_{st} \geq \left(\frac{h(x_{st} + y_{st})}{\sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right)} \right)^{1/\gamma_{s,t}} \implies z_{st} = y_{st} \tag{48}$$

Therefore,

$$B^H_{st}(x_{st}, y_{st}) = \left(\frac{h(x_{st} + y_{st})}{\sqrt{2}\sigma \text{erf}^{-1}\left(2\kappa(\mathbf{c}) - 1\right)} \right)^{1/\gamma_{s,t}} \tag{49}$$

Proof of Proposition 2

To show that $MSPP^{'}$ is discrete convex in (x_{st}, y_{st}), it is sufficient to show that the $MSPP^{'}$ is convex for continuous (x_{st}, y_{st}). In effect, we need to show the integer relaxation of $MSPP'$ is convex in (x_{st}, y_{st}). We start by proving that $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ is convex in (x_{st}, y_{st}) $\forall B_{st} \in [0, \infty)$. For conciseness we drop the subscripts s, t and we represent $\mathcal{W}^{LP}(x, y; \mathbf{c}, B, n)$ by $\mathcal{W}^{LP}(x, y)$. Also, we assume there be (x^1_{st}, y^1_{st}), (x^2_{st}, y^2_{st}), such there exist optimal solutions z^1 and z^2 to $\mathcal{W}^{LP}(x^1, y^1)$ and $\mathcal{W}^{LP}(x^2, y^2)$ respectively. Let $\lambda \in [0, 1]$, and let x^λ, y^λ, z^λ be defined as:

$$x^\lambda_{st} = \lambda x^1_{st} + (1-\lambda) x^2_{st} \tag{50}$$

$$y^\lambda_{st} = \lambda y^1_{st} + (1-\lambda) y^2_{st} \tag{51}$$

$$z^\lambda_{st} = \lambda z^1_{st} + (1-\lambda) z^2_{st} \tag{52}$$

Next, define $\mathcal{W}^{LP}(x^\lambda, y^\lambda)$ as:

$$\mathcal{W}^{LP}(x^\lambda, y^\lambda) = \min_{z \leq y^\lambda} \left\{ c_q z + c_q^{'}(y^\lambda - z) + \mathbf{E}\left[c_o(D - h(x^\lambda + z))^+ + c_u(h(x^\lambda + z) - D)^+ \right] \right\} \tag{53}$$

Note that z^λ is a feasible solution to the mathematical program in (53) as:

$$z^1 \leq y^1 \text{ and } z^2 \leq y^2 \implies \lambda z^1 + (1-\lambda) z^2 \leq \lambda y^1 + (1-\lambda) y^2 \implies z^\lambda \leq y^\lambda \tag{54}$$

Thus,

$$\mathcal{W}^{LP}(x^\lambda, y^\lambda) \le c_q z^\lambda + c_q^{'}(y^\lambda - z^\lambda) + \mathbf{E}\Big[c_o(D - h(x^\lambda + z^\lambda))^+ + c_u(h(x^\lambda + z^\lambda) - D)^+\Big]. \tag{55}$$

We substitute for x^λ and z^λ inside the expectation and simplify to get,

$$\begin{aligned}\mathcal{W}^{LP}(x^\lambda, y^\lambda) \le\, & c_q z^\lambda + c_q^{'}(y^\lambda - z^\lambda) \\ & + \mathbf{E}\Big[c_o(\lambda(D - h(x^1 + z^1)) + (1-\lambda)(D - h(x^2 + z^2)))^+ \\ & + c_u(\lambda(h(x^1 + z^1) - D) + (1-\lambda)(h(x^2 + z^2)) - D)^+\Big]. \end{aligned} \tag{56}$$

As $\max\{0, a+b\} \le \max\{0, a\} + \max\{0, b\}$, we can simplify the above inequality as:

$$\begin{aligned}\mathcal{W}^{LP}(x^\lambda, y^\lambda) \le\, & c_q z^\lambda + c_q^{'}(y^\lambda - z^\lambda) \\ & + \mathbf{E}\Big[\lambda c_o(D - h(x^1 + z^1))^+ + (1-\lambda)c_o(D - h(x^1 + z^1))^+\Big] \\ & + \mathbf{E}\Big[\lambda c_u(h(x^1 + z^1) - D)^+ + (1-\lambda)c_u(D - h(x^1 + z^1))^+\Big]. \end{aligned} \tag{57}$$

Substituting for z^λ and y^λ and collecting terms together, we get:

$$\begin{aligned}\mathcal{W}^{LP}(x^\lambda, y^\lambda) \le\, & \lambda\Big\{c_q z^1 + c_q'(y^1 - z^1) + \mathbf{E}\Big[c_o(D - h(x^1 + z^1))^+ + c_u(h(x^1 + z^1) - D)^+\Big]\Big\} \\ & (1-\lambda)\Big\{c_q z^2 + c_q'(y^2 - z^2) + \mathbf{E}\Big[c_o(D - h(x^2 + z^2))^+ + c_u(h(x^2 + z^2) - D)^+\Big]\Big\}. \end{aligned} \tag{58}$$

Since z^1 and z^2 are optimal solutions for $\mathcal{W}^{LP}(x^1, y^1)$ and $\mathcal{W}^{LP}(x^2, y^2)$ respectively, we can simplify the above as:

$$\mathcal{W}^{LP}(x^\lambda, y^\lambda) \le \lambda \mathcal{W}^{LP}(x^1, y^1) + (1-\lambda)\mathcal{W}^{LP}(x^2, y^2). \tag{59}$$

Therefore, $\mathcal{W}^{LP}(x, y)$ is convex in (x, y).

Since $\mathcal{W}^{LP}(x_{st}, y_{st}; \mathbf{c}, B_{st}, n_{st})$ is convex in (x_{st}, y_{st}) and the expectation operator preserves convexity $\forall B_{st} \in [0, \infty)$, the $MSPP^{'}$ is convex in (x_{st}, y_{st}). This in turn implies that the $MSPP^{'}$ is also discrete convex in (x_{st}, y_{st}) $\forall B_{st} \in [0, \infty)$.

Proof of Proposition 3

From equations (25) and (26) we have:

$$\log \mathcal{L}(\mathbf{c}) = \sum_{s,t} \left\{ -\mathcal{U}(\mathbf{c}, z_{st}, y_{st} B_{st}) - \log\left[\sum_{\tilde{z}_{st} \le \tilde{y}_{st}} \exp\left(-\mathcal{U}(\mathbf{c}, z_{st}, y_{st} B_{st})\right)\right]\right\} \tag{60}$$

To prove that $\log \mathcal{L}(\mathbf{c})$ is concave in $\mathbf{c}$, we prove that $-\mathcal{U}(\mathbf{c}, z_{st}, y_{st}, B_{st}) - \log\left[\sum_{\tilde{z}_{st} \le \tilde{y}_{st}} \exp\left(-\mathcal{U}(\mathbf{c}, z_{st}, y_{st}, B_{st})\right)\right]$ is concave in $\mathbf{c}$ $\forall s, t$.

We first make the transformations $w_z = -\mathcal{U}(\mathbf{c}, z_{st}, y_{st}, B_{st})$ and $\mathbf{w} = (w_1, \ldots, w_{\tilde{y}})$. Let:

$$f_{st}(\mathbf{w}) = w_z - \log\left[\sum_{\tilde{z} \le \tilde{y}} \exp\left(w_z\right)\right]. \tag{61}$$

Then, the Hessian of $f(\mathbf{w})$ is:

$$\nabla^2 f_{st}(\mathbf{w}) = -\nabla^2 \log\left[\sum_{\tilde{z} \le \tilde{y}} \exp\left(w_z\right)\right], \tag{62}$$

where $\log \left[\sum_{\tilde{z}\leq\tilde{y}} \exp\left(w_z\right)\right]$ is the log-sum-exp function. The Hessian of the log-sum-exp function is positive semidefinite (Boyd and Vandenberghe (2004) p. 74). Therefore, from equation (62) the Hessian of $f_{st}(\mathbf{w})$ is negative semidefinite and thus, $f_{st}(\mathbf{w})$ is concave in $\mathbf{w}$. From the definition of $\mathcal{U}(\mathbf{c}, z_{st}, y_{st}B_{st})$ in equation (24), we can see that $\mathcal{U}(\mathbf{c}, z_{st}, y_{st}B_{st})$ is a linear function of $\mathbf{c}$. Therefore, w_z is a linear transform in $\mathbf{c}$. This implies that $f_{st}(\mathbf{w})$ is also concave in $\mathbf{c}$. Since the sum of concave functions is also concave, $\log \mathcal{L}(\mathbf{c}) = \sum_{st} f_{st}(\mathbf{w})$ is concave in $\mathbf{c}$.

References

Aguirregabiria V (1999) The Dynamics of Markups and Inventories in Retailing Firms. *Rev. Econ. Stud.* 66(2):275–308.

Aguirregabiria V, Mira P (2010) Dynamic discrete choice structural models: A survey. *J. Econom.* 156(1):38–67.

Aksin Z, Ata B, Emadi S, Su CL (2013) Structural Estimation of Callers ' Delay Sensitivity in Call Centers. *Management Sci.* 59(12):1–21.

Aksin Z, Ata B, Emadi SM, Su C (2017) Impact of Delay Announcements in Call Centers: An Empirical Approach. *Oper. Res.* 65(1):242–265.

Allenby GM, Arora N, Ginter JL (1998) On the heterogeneity of demand. *J. Mark. Res.* 35(3):384–389.

Allon G, Federgruen A, Pierson M (2011) How Much Is A Reduction Of Your Customers' Wait Worth? An Empirical Study of the Fast-Food Drive-Thru Industry Based On Structural Estimation Methods. *Manufacturing Service Oper. Management* 13(4):489–507.

Bansal S, Nagarajan M (2017) Product portfolio management with production flexibility in agribusiness. *Oper. Res.* 65(4):914–930.

Bard JF, Purnomo HW (2005) Hospital-wide reactive scheduling of nurses with preference considerations. *IIE Trans.* 37(7):589–608.

Bernstein E, Kesavan S, Staats (2014) How to Manage Scheduling. *Harv. Bus. Rev.* .

Birge JR (1985) Decomposition and partitioning methods for multistage stochastic linear programs. *Oper. Res.* 33(5):989–1007.

Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and students as newsvendors. *Management Sci.* 58(12):2225–2233.

Bostian AA, Holt CA, Smith AM (2008) Newsvendor pull-to-center effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management* 10(4):590–608.

Boyd S, Vandenberghe L (2004) *Convex optimization* (Cambridge university press).

Boyd S, Xiao L, Mutapcic A (2003) Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter* 2004:2004–2005.

Brunner JO, Bard JF, Kolisch R (2009) Flexible shift scheduling of physicians. *Health Care Manag. Sci.* 12(3):285–305.

Bureau of Labor Statistics (2018) Highest paid occupations. URL `https://www.bls.gov/ooh/highest-paying.htm`, [Online; accessed 13-August-2018].

Cameron AC, Trivedi PK (2005) *Microeconometrics: methods and applications* (Cambridge University Press).

Chen Y, Song Y (2019) Quantal theory in operations management. *Decision-making in Humanitarian Operations*, 169–191 (Springer).

Deshpande V, Arıkan M (2012) The impact of airline flight schedules on flight delays. *Manufacturing Service Oper. Management* 14(3):423–440.

Dexter F, Ledolter J, Wachtel RE (2005) Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesth. Analg.* 100(5):1425–1432.

Dexter F, O'Neill L (2001) Weekend operating room on call staffing requirements. *AORN J.* 74(5):664–671.

Duan N, Manning WG, Morris CN, Newhouse JP (1983) A Comparison of Alternative Models for the Demand for Medical Care. *J. Bus. Econ. Stat.* 1(2):115–126.

Fisher M, Rajaram K, Raman A (2001) Optimizing Inventory Replenishment of Retail Fashion Products. *Manufacturing Service Oper. Management* 3(3):230–241.

Foreman J, Gallien J, Alspaugh J, Lopez F, Bhatnagar R, Teo CC, Dubois C (2010) Implementing supply-routing optimization in a make-to-order manufacturing network. *Manufacturing & Service Operations Management* 12(4):547–568.

Gade D, Küçükyavuz S, Sen S (2014) Decomposition algorithms with parametric gomory cuts for two-stage stochastic integer programs. *Math. Programming* 144(1-2):39–64.

Gander P, Purnell H, Garden A, Woodward A (2007) Work patterns and fatigue-related risk among junior doctors. *Occupational and environmental medicine* 64(11):733–738.

Greene WH (2003) *Econometric analysis* (Pearson Education), 7th edition.

Grün B, Leisch F (2007) Fitting finite mixtures of generalized linear regressions in R. *Comput. Stat. Data Anal.* 51(11):5247–5252.

Gurnani H, Tang CS (1999) Note: Optimal Ordering Decisions with Uncertain Cost and Demand Forecast Updating. *Management Sci.* 45(10):1456–1462.

Haufler K, Harrington M (2011) Using nurse-to-patient telephone calls to reduce day-of-surgery cancellations. *AORN J.* 94(1):19–26.

He B, Dexter F, Macario A, Zenios S (2012) The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem. *Manufacturing Service Oper. Management* 14(1):99–114.

Healthcare Insights (2014) The evidence is clear: Analytics key to controlling labor costs inefficiency is no longer an option. *White Paper* .

Henly JR, Lambert SJ (2014) Unpredictable work timing in retail jobs: Implications for employee work–life conflict. *ILR Review* 67(3):986–1016.

Ho TH, Lim N, Cui TH (2010) Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science* 56(11):1891–1910.

Johnson S (2014) The NLOPT nonlinear-optimization package [software]. `http://ab-initio.mit.edu/nlopt`.

Kantor J (2014) Working anything but 9 to 5: scheduling technology leaves low-income parents with hours of chaos. *The New York Times* .

Kantor J (2015) Starbucks to revise policies to end irregular schedules for its 130,000 baristas. *The New York Times* .

Keehan SP, Stone DA, Poisal JA, Cuckler GA, Sisko AM, Smith SD, Madison AJ, Wolfe CJ, Lizonitz JM (2017) National health expenditure projections, 2016-25: Price increases, aging push sector to 20 percent of economy. *Health Aff.* 36(3):553–563.

Kesavan S, Staats BR, Gilland W (2014) Volume Flexibility in Services : The Costs and Benefits of Flexible Labor Resources Volume Flexibility in Services . *Management Sci.* 60(8):1884–1906.

Knox M, Myers E, Wilson I, Hurley M (2009) The impact of pre-operative assessment clinics on elective surgical case cancellations. *The Surgeon* 7(2):76–78.

Kohavi R, et al. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *ICJAI*.

Kong N, Schaefer AJ, Ahmed S (2013) Totally unimodular stochastic programs. *Math. Programming* 138(1-2):1–13.

May JH, Strum DP, Vargas LG (2000) Fitting the lognormal distribution to surgical procedure times. *Decision Sciences* 31(1):129–148.

McFadden DL (1974) Analysis of qualitative choice behavior. Zaremka P, ed., *Frontiers in Econometrics* (Academic Press. New York, NY).

McFadden DL (1976) Quantal choice analaysis: A survey. *Annals of Economic and Social Measurement, Volume 5, number 4*, 363–390 (NBER).

McFadden DL, Manski CF (1981) *Econometric models of probabilistic choice* (MIT Press).

McIntosh C, Dexter F, Epstein RH (2006) The Impact of Service-Specific Staffing, Case Scheduling, Turnovers, and First-Case Starts on Anesthesia Group and Operating Room Productivity: A Tutorial Using Data from an Australian Hospital. *Anesth. Analg.* 103:1499–516.

Milne RG, Horne M, Torsney B (2006) Sms reminders in the uk national health service: an evaluation of its impact on" no-shows" at hospital out-patient clinics. *Health Care Manage. Rev.* 31(2):130–136.

Min Y, Agresti A (2002) Modeling Nonnegative Data with Clumping at Zero : A Survey Models for Semi-continuous Data. *JIRSS* 1(May):7–33.

Nahmias S, Cheng Y (2009) *Production and operations analysis*, volume 4 (McGraw-Hill/Irwin New York).

Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Sci.* 54(1):41–55.

Pinker EJ, Larson RC (2003) Optimizing the use of contingent labor when demand is uncertain. *Eur. J. Oper. Res.* 144(1):39–55.

Rogers AE, Hwang WT, Scott LD, Aiken LH, Dinges DF (2004) The working hours of hospital staff nurses and patient safety. *Health Aff.* 23(4):202–212.

Rust J (1987) Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica* 999–1033.

Schultz R, Stougie L, Van Der Vlerk MH (1998) Solving stochastic programs with integer recourse by enumeration: A framework using gröbner basis. *Mathematical Programming* 83(1-3):229–252.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* 46(3):404–420.

Smith M, Saunders R, Stuckhardt L, Mcginnis JM (2012) *Best Care at Lower Cost The Path to Continuously Learning Health Care in America Committee on the Learning Health Care System in America* (National Academic Press).

Stimpfel AW, Sloane DM, Aiken LH (2012) The longer the shifts for hospital nurses, the higher the levels of burnout and patient dissatisfaction. *Health Aff.* 31(11):2501–2509.

Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure timescomparison of log-normal and normal models. *Anesthesiology* 92(4):1160–1167.

Su X (2008) Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4):566–589.

Sun RR, Shylo OV, Schaefer AJ (2015) Totally unimodular multistage stochastic programs. *Oper. Res. Lett.* 43(1):29–33.

Tiwari V, Furman WR, Sandberg WS (2014) Predicting case volume from the accumulating elective operating room schedule facilitates staffing improvements. *Anesthesiology* 121(1):171–183.

Trinkoff AM, Le R, Geiger-Brown J, Lipscomb J, Lang G (2006) Longitudinal relationship of work hours, mandatory overtime, and on-call to musculoskeletal problems in nurses. *Am. J. Ind. Med.* 49(11):964–971.

Wachtel RE, Dexter F (2010) Review of behavioral operations experimental studies of newsvendor problems for operating room management. *Anesth. Analg.* 110(6):1698–1710.

Wild B, Schneewei C (1993) Manpower capacity planning — A hierarchical approach. *Int. J. Prod. Econ.* 30:95–106.

Xie S, Zenios SA (2015) Forecasting and dynamic adjustment of staffing levels in hospital operating rooms. Working paper, Stanford Graduate School of Business.

Tables and Figures

Table 1 Summary statistics for historical anesthesiologist planning by specialty

Specialty	Staff type	Average	Max	Min
Cardio-Thoracic	Regular	4.93	10	0
	On-call consideration	1.18	6	0
	On-call actually called	0.45	5	0
General	Regular	8.65	16	0
	On-call consideration	4.61	7	0
	On-call actually called	1.85	10	0
Neuro	Regular	2.72	6	0
	On-call consideration	0.72	4	0
	On-call actually called	0.30	3	0
Pediatric	Regular	1.69	6	0
	On-call consideration	0.53	4	0
	On-call actually called	0.24	4	0
Total	Regular	17.48	26	0
	On-call consideration	6.89	11	0
	On-call actually called	2.77	9	0

Table 2 Maximum Likelihood Estimates of Implicit Cost Parameters

Cost Parameters	Maximum Likelihood Estimate*	95% Confidence Intervals (Bootstrap)
$c_q^{'}$	1.56	$(1.02, 2.09)$
c_u	0.35	$(0.21, 0.49)$

*Values scaled such that $c_q = 1$

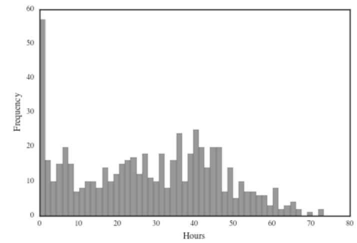
Table 3 Daily Average Percent Cost Saving of Model Based Solution Over Current Practice

Specialty	Daily Average Cost Saving (%)
Cardio-Thoracic	5.12
General	14.44
Neuro	24.76
Pediatric	16.24
Average	**13.72**

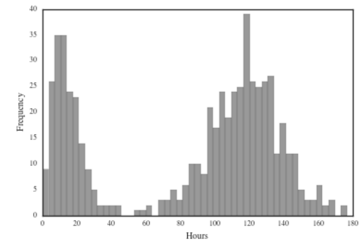
Table 4 Comparison of Model Based Staffing Plan to Current Practice

	Model Based Staffing Plan	Current Practice
Average daily overtime (hours)	54.06	67.76
Average daily idle time (hours)	18.53	29.36
Average number of anesthesiologists on regular duty	17.08	17.48
Average number of anesthesiologists on on-call consideration list	7.28	6.89
Average number of anesthesiologists called	3.49	2.77
Percentage of days with no on-call consideration list	47.29%	31.35%

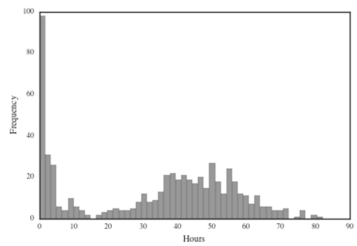
Figure 1 **Distribution of surgery booked time by specialty**



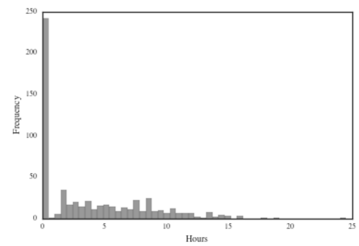
(a) Cardio-Thoracic



(b) General



(c) Neuro



(d) Pediatric

Figure 2 **Impact of change in cost parameters**

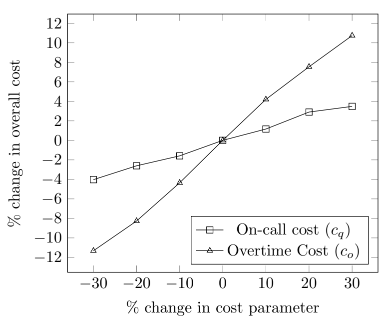


Figure 3 **Impact of change in standard deviation of booked time (B_{st})**

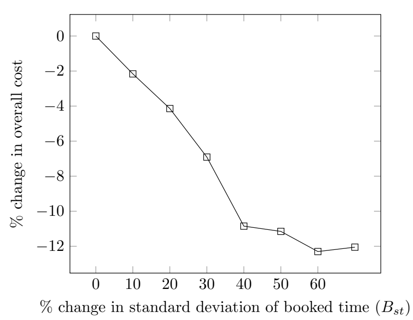
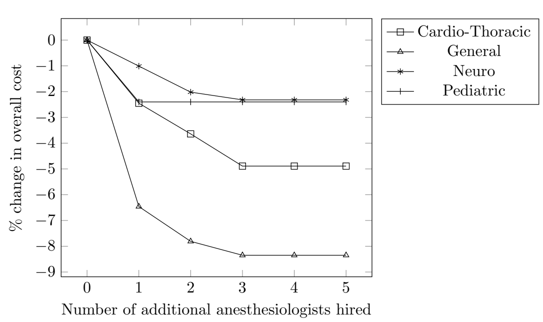


Figure 4 **Impact of hiring additional anesthesiologists**



Electronic Companion
Staff Planning for Hospitals with Cost Estimation and Optimization

Sandeep Rath
Kenan-Flagler Business School, University of North Carolina at Chapel Hill
sandeep@unc.edu
Kumar Rajaram
UCLA Anderson School of Management, Los Angeles CA 90095
kumar.rajaram@anderson.ucla.edu

EC.1. Estimation Results for Distribution of Booked Hours by Specialty (B_{st})

1. Cardio-Thoracic Surgeries

Table EC.1 Logit Regression for Cardio-Thoracic Surgeries $B_{s,t} = 0$

Dep. Variable:	CARDIO-THORACIC	**No. Observations:**	591		
Model:	Logit	**Df Residuals:**	587		
Method:	MLE	**Df Model:**	3		
		Pseudo R-squ.:	0.2702		
		Log-Likelihood:	-133.51		
		LL-Null:	-182.95		
	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
intercept	3.1611	0.248	12.759	0.000	2.676 3.647
Monday	3.6758	1.455	2.526	0.012	0.824 6.528
Sunday	-2.7066	0.333	-8.127	0.000	-3.359 -2.054
True	-4.6557	1.126	-4.136	0.000	-6.862 -2.450

Table EC.2 Log Normal Regression for Cardio-Thoracic Surgeries B_{st}

Dep. Variable:	CARDIO-THORACIC	**R-squared:**	0.677
Model:	OLS	**Adj. R-squared:**	0.668
Method:	Least Squares	**F-statistic:**	72.80
Date:	Mon, 18 Sep 2017	**Prob (F-statistic):**	4.12e-117
Time:	17:22:37	**Log-Likelihood:**	-356.06
No. Observations:	536	**AIC:**	744.1
Df Residuals:	520	**BIC:**	812.7
Df Model:	15		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Wednesday	-0.3104	0.061	-5.087	0.000	-0.430 -0.190
Friday	-0.3476	0.060	-5.760	0.000	-0.466 -0.229
Saturday	-1.6969	0.066	-25.835	0.000	-1.826 -1.568
Sunday	-1.6141	0.073	-21.985	0.000	-1.758 -1.470
February	0.4576	0.115	3.970	0.000	0.231 0.684
March	0.3712	0.090	4.141	0.000	0.195 0.547
April	0.2721	0.092	2.971	0.003	0.092 0.452
May	0.3816	0.092	4.170	0.000	0.202 0.561
June	0.4020	0.092	4.393	0.000	0.222 0.582
July	0.5096	0.092	5.565	0.000	0.330 0.690
August	0.5706	0.092	6.233	0.000	0.391 0.750
September	0.3650	0.090	4.058	0.000	0.188 0.542
October	0.6102	0.097	6.269	0.000	0.419 0.801
November	0.3453	0.114	3.036	0.003	0.122 0.569
True	-1.4289	0.163	-8.771	0.000	-1.749 -1.109
intercept	3.3031	0.068	48.612	0.000	3.170 3.437

2. Neuro Surgeries

Table EC.3 Logit Regression for Neuro Surgeries $B_{st} = 0$

Dep. Variable:	NEURO	**No. Observations:**	591
Model:	Logit	**Df Residuals:**	587
Method:	MLE	**Df Model:**	3
		Pseudo R-squ.:	0.4526
		Log-Likelihood:	-142.64
		LL-Null:	-260.57

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
intercept	4.9031	0.580	8.460	0.000	3.767 6.039
Saturday	-4.7382	0.619	-7.653	0.000	-5.952 -3.525
Sunday	-5.2113	0.620	-8.409	0.000	-6.426 -3.997
True	-3.9868	0.828	-4.814	0.000	-5.610 -2.364

Table EC.4 Log Normal Regression for Neuro Surgeries B_{st}

Dep. Variable:	NEURO	**R-squared:**	0.820
Model:	OLS	**Adj. R-squared:**	0.818
Method:	Least Squares	**F-statistic:**	372.3
		Prob (F-statistic):	9.69e-179
		Log-Likelihood:	-268.48
No. Observations:	496	**AIC:**	551.0
Df Residuals:	489	**BIC:**	580.4
Df Model:	6		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Wednesday	-0.1799	0.053	-3.367	0.001	-0.285 -0.075
Friday	-0.1660	0.053	-3.114	0.002	-0.271 -0.061
Saturday	-2.3148	0.067	-34.357	0.000	-2.447 -2.182
Sunday	-2.4899	0.075	-33.276	0.000	-2.637 -2.343
October	0.1910	0.071	2.695	0.007	0.052 0.330
True	-1.8760	0.135	-13.896	0.000	-2.141 -1.611
intercept	3.8456	0.027	140.079	0.000	3.792 3.900

3. Pediatric Surgeries

Table EC.5 Logit Regression for Pediatric Surgeries $B_{st} = 0$

Dep. Variable:	PEDS	**No. Observations:**	591
Model:	Logit	**Df Residuals:**	583
Method:	MLE	**Df Model:**	7
		Pseudo R-squ.:	0.3025
		Log-Likelihood:	-278.94
converged:	True	**LL-Null:**	-399.91

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
intercept	2.0440	0.205	9.962	0.000	1.642 2.446
Monday	-1.2851	0.316	-4.069	0.000	-1.904 -0.666
Saturday	-3.0456	0.317	-9.603	0.000	-3.667 -2.424
Sunday	-3.8254	0.362	-10.555	0.000	-4.536 -3.115
Tuesday	-2.6601	0.307	-8.664	0.000	-3.262 -2.058
December	1.0096	0.483	2.092	0.036	0.064 1.955
October	0.8964	0.423	2.119	0.034	0.067 1.725
True	-2.4334	0.651	-3.741	0.000	-3.708 -1.158

Table EC.6 Log Normal Regression for Pediatric Surgeries B_{st}

Dep. Variable:	PEDS	**R-squared:**	0.279
Model:	OLS	**Adj. R-squared:**	0.271
Method:	Least Squares	**F-statistic:**	33.27
		Prob (F-statistic):	1.82e-23
		Log-Likelihood:	-306.77
No. Observations:	349	**AIC:**	623.5
Df Residuals:	344	**BIC:**	642.8
Df Model:	4		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Tuesday	-0.6983	0.112	-6.252	0.000	-0.918 -0.479
Thursday	0.1670	0.079	2.115	0.035	0.012 0.322
Saturday	-0.8640	0.124	-6.942	0.000	-1.109 -0.619
Sunday	-1.0534	0.162	-6.493	0.000	-1.372 -0.734
intercept	1.7966	0.041	43.501	0.000	1.715 1.878

4. General Surgeries

Table EC.7 First Component for Gaussian Mixture Regression for general Surgeries B_{st}

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	4.696350	0.036904	127.2578	$< 2.2e-16$ ***
Monday	-0.082861	0.031581	-2.6238	0.008696 **
Saturday	-2.074641	0.065896	-31.4835	$< 2.2e-16$ ***
Sunday	-2.353900	0.097178	-24.2226	$< 2.2e-16$ ***
Tuesday	0.061990	0.029223	2.1213	0.033898 *
December	0.209230	0.053643	3.9005	$9.601e-05$ ***
January	0.112182	0.052875	2.1216	0.033869 *
July	0.101713	0.044038	2.3097	0.020907 *
June	0.083648	0.042617	1.9628	0.049669 *
November	0.216889	0.054965	3.9460	$7.948e-05$ ***
September	0.103088	0.044365	2.3236	0.020145 *
holiday	-1.466390	0.111126	-13.1958	$< 2.2e-16$ ***

Table EC.8 Second Component for Gaussian Mixture Regression for general Surgeries $B_{s,t}$

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	4.359169	0.239888	18.1717	$< 2.2e-16$ ***
Saturday	-1.531509	0.212058	-7.2221	$5.118e-13$ ***
Sunday	-2.318971	0.209306	-11.0793	$< 2.2e-16$ ***
December	-1.072790	0.289969	-3.6997	0.0002159 ***
holiday	-2.020018	0.236400	-8.5449	$< 2.2e-16$ ***

EC.2. Estimation Results for Distribution Hours by Specialty Conditioned on Booked Hours ($D_{st}|B_{st}$)

1. Cardio-Thoracic Surgeries

Table EC.9 Regression model for $D_{s,t}$ for Cardio-Thoracic surgeries

Dep. Variable:	$D_{s,t}$	**R-squared:**	0.978		
Model:	OLS	**Adj. R-squared:**	0.978		
Method:	Least Squares	**F-statistic:**	2.434e+04		
		Prob (F-statistic):	0.00		
		Log-Likelihood:	-341.43		
No. Observations:	536	**AIC:**	684.9		
Df Residuals:	535	**BIC:**	689.1		
Df Model:	1				
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
$B_{s,t}$	0.9369	0.006	156.000	0.000	0.925 0.949

2. General Surgeries

Table EC.10 Regression model for $D_{s,t}$ for General surgeries

Dep. Variable:	$D_{s,t}$	**R-squared:**	0.991		
Model:	OLS	**Adj. R-squared:**	0.991		
Method:	Least Squares	**F-statistic:**	6.648e+04		
		Prob (F-statistic):	0.00		
		Log-Likelihood:	-280.81		
No. Observations:	587	**AIC:**	563.6		
Df Residuals:	586	**BIC:**	568.0		
Df Model:	1				
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
$B_{s,t}$	0.9896	0.004	257.828	0.000	0.982 0.997

3. Neuro Surgeries

Table EC.11 Regression model for $D_{s,t}$ for Neuro surgeries

Dep. Variable:	$D_{s,t}$	**R-squared:**	0.982		
Model:	OLS	**Adj. R-squared:**	0.982		
Method:	Least Squares	**F-statistic:**	2.714e+04		
		Prob (F-statistic):	0.00		
Time:	11:59:14	**Log-Likelihood:**	-298.17		
No. Observations:	496	**AIC:**	598.3		
Df Residuals:	495	**BIC:**	602.6		
Df Model:	1				
	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
$B_{s,t}$	0.9313	0.006	164.734	0.000	0.920 0.942

4. Pediatric Surgeries

Table EC.12 Regression model for $D_{s,t}$ for Pediatric surgeries

Dep. Variable:	$D_{s,t}$	**R-squared:**	0.982
Model:	OLS	**Adj. R-squared:**	0.982
Method:	Least Squares	**F-statistic:**	2.714e+04
		Prob (F-statistic):	0.00
		Log-Likelihood:	-298.17
No. Observations:	496	**AIC:**	598.3
Df Residuals:	495	**BIC:**	602.6
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
$B_{s,t}$	0.9313	0.006	164.734	0.000	0.920 0.942

EC.3. Computational Analysis for Performance of Proposed Heuristic Method

In Table EC.13 we compute the relative gap given by:

$$\text{Percentage Relative Gap} = \frac{\text{Model Based Solution Objective } - \text{Complete Enumeration Solution Objective}}{\text{Complete Enumeration Solution Objective}} \times 100\% \tag{EC.1}$$

We compute the percentage relative gap for different problems sizes. Each problem consists of all $S = \{Cardio - Thoracic, General, Neuro, Pediatric\}$ specializations and T days as given in Table EC.13 below.

Table EC.13 Relative gap and solution time of Model based solution and complete enumeration based solution

Problem Size (T)	Percentage Relative Gap	Computation time of enumeration based method (in hours)	Computation time of model based solution (in hours)
1	0.53	0.87	0.01
5	0.75	3.25	0.01
10	1.23	14.37	0.06
20	1.88	37.33	0.08
30	1.94	98.24	0.15

EC.4. Estimation and Validation Results for Model with Different Implicit Costs for Each Specialty

Table EC.14 Maximum Likelihood Estimates of Implicit Cost Parameters

Specialty (s)	$c_q^{'}$	c_u
Cardio-Thoracic	1.46 (0.87-2.04)	0.32 (0.19-0.44)
General	1.45 (0.75-2.16)	0.29 (0.11-0.46)
Neuro	1.36 (0.94-1.76)	0.28 (0.15-0.42)
Pediatric	1.78 (1.70-1.84)	0.44 (0.36-0.51)
Average RMSE: 1.59		
Average Accuracy: 37.46%		

Numbers in brackets indicate 95% confidence interval

EC.5. Estimation Results for Model with Different Implicit Costs for Two Halves of the Data Set

Table EC.15 Maximum Likelihood Estimates of Implicit Cost Parameters

	$c_q^{'}$	c_u
First half of data set	1.63 (1.35, 1.99)	0.33 (0.301, 0.37)
Second half of data set	1.75 (1.44, 1.85)	0.37 (0.31, 0.44)

Numbers in brackets indicate 95% confidence interval

EC.6. Estimation Results for Model with Different Implicit Costs for Each Day of Week

Table EC.16 Maximum Likelihood Estimates of Implicit Cost Parameters by day of week

	$c_q^{'}$	c_u
Monday	1.77 (1.46, 2.07)	0.41 (0.27, 0.53)
Tuesday	1.58 (1.30, 1.86)	0.37 (0.28, 0.54)
Wednesday	1.75 (1.67, 1.85)	0.34 (0.26, 0.41)
Thursday	1.56 (1.33, 1.78)	0.39 (0.3, 0.49)
Friday	1.65 (1.29, 2.01)	0.29 (0.14, 0.37)

Numbers in brackets indicate 95% confidence interval

EC.7. Modeling Decision Model of Staff Planner as a Linear Decision Rule

Regression results assuming the staff planner follows a linear decision rule, as given below:

$$z_{st} = B_{st} + x_{st} + y_{st} + e \tag{EC.2}$$

Table EC.17 Regression model for Linear Decision Rule

Dep. Variable:	z	**R-squared:**	0.638
Model:	OLS	**Adj. R-squared:**	0.636
Method:	Least Squares	**F-statistic:**	323.1
		Prob (F-statistic):	8.20e-121
		Log-Likelihood:	-720.04
No. Observations:	552	**AIC:**	1446.
Df Residuals:	549	**BIC:**	1459.
Df Model:	3		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
$B_{s,t}$	0.0578	0.024	2.438	0.015	0.011 0.104
$x_{s,t}$	-0.0785	0.017	-4.602	0.000	-0.112 -0.045
$y_{s,t}$	0.4917	0.023	21.065	0.000	0.446 0.538

Average RMSE for k-fold out of sample prediction = 0.89

EC.8. Evidence of Mean Anchoring Bias

In this section, we test if the staff planner demonstrates mean anchoring bias. We follow Bostian et al. (2008), to show that if the staff planner demonstrates mean anchoring bias, then the observed number of anesthesiologists actually called (z_{st}^{obs}) will be given by:

$$z_{st}^{obs} - \bar{z}_{st} = \alpha(z_{st}^{opt} - \bar{z}_{st}) + \epsilon' \tag{EC.3}$$

Here, z_{st}^{opt} is the optimal number of anesthesiologists to call, obtained by applying Proposition 1, $\bar{z}_{st}$ is the number of anesthesiologists to call to meet the mean demand. The mean demand, as described in Section 3.2 would be $\bar{D}_{st} = \gamma B_{st}$. Since there are x_{st} anesthesiologists already on regular duty, the number of anesthesiologists to call to meet the mean demand would therefore be, $\bar{z}_{st} = (\bar{D}_{st} - hx_{st})^+/h$ rounded up to the nearest integer value.

If we regress $z_{st}^{obs} - \bar{z}_{st}$ on $z_{st}^{opt} - \bar{z}_{st}$, we get an estimate for α. As described in Bostian (2008), if $\alpha \in (0,1)$, then the decision of the staff planner would be consistent with mean anchoring. We report the results of the regression below and we can observe that the estimates of α is in $(0,1)$ which would be consistent with mean anchoring of decisions by the staff planner.

Table EC.18 Regression model for Linear Decision Rule

Dep. Variable:	$z_{st}^{obs} - \bar{z}_{st}$	**R-squared:**	0.824
Model:	OLS	**Adj. R-squared:**	0.824
Method:	Least Squares	**F-statistic:**	2582.
		Prob (F-statistic):	4.09e-210
		Log-Likelihood:	-748.24
No. Observations:	552	**AIC:**	1498.
Df Residuals:	551	**BIC:**	1503.
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
α	0.6622	0.013	50.816	0.000	0.637 0.688

References

Bostian, AJ A., Charles A. Holt, and Angela M. Smith. "Newsvendor pull-to-center effect: Adaptive learning in a laboratory experiment." Manufacturing & Service Operations Management 10.4 (2008): 590-608.