Integrated Anesthesiologist and Room Scheduling for Surgeries: Methodology and Application

Sandeep Rath

Kenan-Flagler Business School
University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
sandeep@unc.edu

Kumar Rajaram

UCLA Anderson School of Management, Los Angeles CA 90095
kumar.rajaram@anderson.ucla.edu

Aman Mahajan

Department of Anesthesiology and Perioperative Medicine
David Geffen School of Medicine at UCLA, Los Angeles CA 90095
amahajan@mednet.ucla.edu

We consider the problem of minimizing daily expected resource usage and overtime costs across multiple parallel resources such as anesthesiologists and operating rooms, which are used to conduct a variety of surgical procedures at large multi-specialty hospitals. To address this problem, we develop a two-stage mixed-integer stochastic dynamic programming model with recourse. The first stage allocates these resources across multiple surgeries with uncertain durations and prescribes the sequence of surgeries to these resources. The second stage determines actual start times to surgeries based on realized durations of preceding surgeries and assigns overtime to resources to ensure all surgeries are completed using the allocation and sequence determined in the first stage. We develop a data driven robust optimization method that solves large-scale real-sized versions of this model close to optimality. We validate and implement this model as a decision support system at the UCLA Ronald Reagan Medical Center. This system effectively incorporates the flexibility in the resources and uncertainty in surgical durations, and explicitly trades off resource usage and overtime costs. This has increased the average daily utilization of the anesthesiologists by 3.5% and of the operating rooms by 3.8%. This has led to an average daily cost savings of around 7% or estimated to be $2.2 million on an annual basis. In addition, the insights based on this model have significantly influenced decision making at the operating services department at this hospital.

1. Introduction

Surgical procedures are complex tasks requiring the use of several specialized and expensive resources. In a hospital, the operating services department is responsible for managing resources used in surgical procedures. Every day, this department assigns to each surgery an operating room, an anesthesiologist, a nursing team and the requisite surgical materials. The department also determines the sequence in which these surgeries will be performed and the scheduled start times. While performing these actions, the department ensures that the cost of the operating room suite is minimized by reducing resource usage and overtime costs.

The operating services departments at large hospitals devote significant amount of time in making these resource management decisions. The complexity of these decisions is due to the following four primary reasons. First, operating room resources are expensive (Macario 2010), and in short supply (Orkin et al. 2013) and thus, surgeries are performed in highly resource constrained environments.

Second, surgical procedures are often very specialized. Therefore, equipment and facility requirements govern whether a procedure can be performed in a particular room. Anesthesiologist assignments too are dictated by specialty. Studies have demonstrated that not only do surgeons often prefer to have an anesthesiologist of the required sub-specialty (Ghaly 2014), outcome indicators of surgical procedures are significantly better when anesthesia is delivered by an anesthesiologist with experience in that particular sub-specialty (McNicol 1997). Pardo (2014) predict that increasingly anesthesiologists will be assigned by their sub-specialties. Third, the durations of surgical procedures are very difficult to predict (Kayis et al. 2012). This is partly because there is a large number of procedures, and newer procedures are constantly being developed (AMA 2013). Consequently, historical data on all these procedures is not available. Furthermore, surgeon's estimates of durations are often unreliable. Studies have demonstrated systematic underestimation as well as overestimation of procedure times by surgeons while scheduling surgeries. Some surgeons overestimate the duration when they do not have enough cases to fill their scheduled block time, while others may underestimate the time when they wish to fill in more cases (Laskin et al. 2013). Finally, the scale of large hospitals, in terms of the number of operating rooms, procedures conducted, the number and types of equipment and anesthesiologists used, makes the simultaneous scheduling of multiple resources a computationally challenging task.

We were exposed to these complexities at the operating services department of the UCLA Ronald Reagan Medical Center (RRMC), a large multi-specialty hospital which consistently ranks amongst the best five hospitals in the United States (Roxanne Moster 2014). The management of this department felt that the daily resource allocation decision played a significant role in overall department cost, and in the service quality delivered to patients. They believed that these aspects can be significantly improved by developing an analytical model based approach that considered the key complexities in this environment, and applied historical surgical data to decide resource assignment and scheduling. This paper describes the development, implementation, and evaluation of a model based decision support system that uses a data driven robust optimization procedure to determine the daily scheduling of anesthesiologists and rooms for elective surgeries at UCLA RRMC.

There is a large body of literature on elective surgery scheduling. Min and Yih (2010) consider scheduling elective surgeries under uncertainty in surgery durations and downstream capacity constraints. Gupta (2007) discusses the broader issues of managing operating room suites for elective surgeries. In this context, we study the problem of integrated scheduling of anesthesiologists and operating rooms to surgeries. The literature on scheduling anesthesiologists to surgeries includes Marcon and Dexter (2006), McIntosh et al. (2006), Dexter and Wachtel (2014), and Dexter et al. (2016). However, these papers do not consider the joint scheduling with operating rooms and uncertainty in surgical durations, both of which were critical features in our application. Furthermore, the solution methods in these papers employ experience-based heuristics rather than optimization-based methods. The literature for scheduling operating rooms under uncertain surgery durations has primarily been focused on single resource

type corresponding to the operating room. Research in this area include Denton and Gupta (2003), Green and Savin (2008), Mancilla and Storer (2012), and Mak et al. (2014a). In addition, Denton et al. (2010) solve the problem of assignment of surgeries to multiple parallel operating rooms under fixed costs of operating rooms and variable overtime costs. However, none of these papers consider multiple resources, the simultaneous sequencing and start times of surgeries or are tested with data in a large-scale application context. While a single resource type may be sufficient for specialized surgery suites, multi-specialty hospitals like the UCLA RRMC require a holistic solution of surgery scheduling that simultaneously optimizes on all specialized parallel and multiple resources.

The literature on multiple resources can be classified in two broad categories: serial and parallel multiple resources. The research on serial multiple resources focuses on analyzing the impact of decisions made on an upstream resource such as the operating rooms on a downstream resource such as a post anesthesia care unit (Marcon and Dexter 2006, Augusto et al. 2010 and Saadouli et al. 2015). Additional work in this area that consider other upstream and downstream resources include Cardoen et al. (2009a) & (2009b), and Gul et al. (2011). However, none of these papers consider parallel multiple resources (such as anesthesiologists and operating rooms), which are specialized and have to be scheduled simultaneously under uncertainty in surgical durations. These were important aspects in the application and significantly complicated the optimization model. The literature related to parallel multiple resource types is relatively scarce. Beliën and Demeulemeester (2008) and Meskens et al. (2013) consider integrated operating room scheduling with multiple parallel resources under deterministic surgery durations. Batun et al. (2011) consider scheduling of surgeries given two parallel resource types: operating rooms and surgeons under stochastic surgery durations. However, they do not consider specializations of rooms and anesthesiologists, and consider a problem significantly smaller in scale than in our application. For the scale of problem at UCLA RRMC, sample average approximation based stochastic optimization procedures as used in Denton and Gupta (2003) and Min and Yih (2010) were intractable. This is due to the large number of possible integer assignments in the first stage which increases the number of samples required to achieve convergence in objective value and solution. These difficulties in employing this method has been described in a more general context by Kleywegt et al. (2002). Furthermore, the overall complexity of the large-scale problem in our application precluded finding even feasible solutions using leading commercially available solvers such as `ddsip` (Märkert and Gollmer 2008) that employ the state of the art procedures for solving stochastic programs such as dual decomposition methods (Carøe and Schultz 1999). We describe this in our computational analysis. In order to circumvent these problems, we use a robust optimization procedure (Bertsimas and Thiele 2006, Bertsimas et al. 2013). While a similar approach has been used by Denton et al. (2010) and Mak et al. (2014b), our work extends theirs by considering multiple resource types. This extension requires significant modification to existing solution methods.

Our paper makes the following contributions. First, we consider two types of parallel resources which are of critical importance to specialties: operating rooms and anesthesiologists, and we simultaneously

optimize their assignment and sequencing. Second, we develop an efficient solution method using robust optimization to provide effective solutions to large-scale problems. An important element when applying robust optimization is the estimation of an uncertainty set. We develop an estimation procedure to estimate the sets using historical data. This data driven robust optimization approach was successful in solving the full scale problem for the entire surgery suite at the RRMC within 25 minutes, with a performance gap within 5% from the lower bound. Third, our methodology significantly outperforms the best benchmark procedures in the literature. Fourth, we develop a model based decision support system which has been validated and implemented at the UCLA RRMC. To the best of our knowledge, this is the first real implementation of robust optimization. This system effectively incorporates the flexibility in the resources and uncertainty in surgical durations, and explicitly trades off costs. This has considerably improved upon current practice, and has resulted in average daily cost savings of around 7% or estimated to be $2.2 million on an annual basis. Further, the insights from our work has had a notable impact on decision making at the hospital.

The remainder of the paper is organized as follows. §2 provides a detailed problem description at the UCLA RRMC. §3 presents the model formulation, properties and solution procedure. In §4 we describe the procedure for parameter estimation and model calibration. §5 provides the results of the computational analysis. §6 describes the implementation of this model at the UCLA RRMC, presents the financial benefits, provides managerial insights and describes the organizational impact of this work.

2. Problem Description

Operating services is one of the largest departments at the UCLA RRMC with around $120 million in annual revenues representing about 10% of this hospital's revenues. This department serves around 27,000 patients annually by conducting around 2700 *types of* elective and emergency surgical procedures across 12 specialties. Emergency surgeries are conducted in 3 dedicated operating rooms with a separate team of anesthesiologists. Since emergency surgeries are separated from elective surgeries and account for only about 15% of revenues, the management of this department asked us to focus solely on elective surgeries. To perform these surgeries, the operating services department uses 23 Operating Rooms (ORs), which are further divided across these 12 specialties that require specific equipment. The details on the number of rooms that can perform each specialty is provided in Table 1. General surgery procedures can be performed in any of these 23 ORs dedicated for the exclusive use of elective surgeries. Add-on surgeries are not considered here as depending on availability, they are assigned to the operating rooms dedicated to emergency procedures. Surgeries are scheduled to start in operating rooms only between 7 am and 3 pm. Further, there are fixed costs for opening an operating room each day. This consists of an initial cleaning and equipment setup costs along with daily nurse and technician staffing costs, whose assignments do not depend on specialty. In addition, overtime costs are incurred for nurses and technicians if the rooms are required to be open beyond 3 pm. Finally, these operating rooms are scheduled and staffed simultaneously.

Insert Table 1 about here

There were 92 anesthesiologists at the UCLA RRMC divided across these 12 specialties. The number of anesthesiologists by specialty is also shown in Table 1. The assignment of anesthesiologists is according to the specialty required for the surgery. Anesthesia for surgeries in some specialties can be administered by any anesthesiologist. Such specialties are denoted by NA in this table. There are three shifts of equal duration for the anesthesiologists: day (7 a.m till 3 p.m.), late (11 a.m. till 7 p.m), and night (7 p.m till 3 a.m). Each anesthesiologist is preassigned to exactly one shift and thus, the regular working hours for each anesthesiologist is eight hours. In addition, anesthesiologists can only be assigned to surgeries that begin during their shift. Overtime costs for anesthesiologists are incurred if surgeries in progress exceed the duration of the shift. A certain number of anesthesiologists who are not scheduled to work on a given day are asked to be on standby or on call, so that they can be called to work if necessary. However, when anesthesiologists are assigned from call, there are significant costs for using such an option. Anesthesiologists assigned from call do not incur overtime costs. The anesthesiologists on call are informed of their status the previous day and assigned surgeries that day as required.

It is important to note that in the context of large multi-specialty hospitals such as the UCLA RRMC, surgeons are not part of the operating services department. They are usually from the independently administered specialty departments at this hospital and on some occasions can be from other hospitals. The surgeons bring their patients and use the operating services department as a service provider. Thus, the operating services department does not have the option of assigning surgeons to patients. For this reason, we assume that each surgery-surgeon combination is already set and we consider them together. This ensures that each surgery has a clear and unchangeable link to the surgeon. This aspect is also consistent with the literature in this area (Dexter and Traub 2002, Marques et al. 2014).

Typically, a request to schedule a surgery is initiated by the surgeon on behalf of the patient with general admissions at the hospital. This request is assigned a date based on the earliest availability in the block reservations for the particular specialty. Once all the elective surgery requests have been received the day before the surgery, the operating services department decides which operating room to open, finalizes assignment of these rooms and anesthesiologist to surgeries, determines start times of surgeries and effectively specifies the sequence of all the surgeries. These decisions are made in the previous day for all the surgeries that need to be conducted in the following day. Consequently, the planning horizon is a single day. The current planning process to make these decisions uses an experience-based practitioner's heuristic. Such types of heuristics have been reported in the literature (Dexter and Traub 2002, Cardoen et al. 2010). The practitioner's heuristic consists of the following steps.

Step 1: Assign surgeries to operating rooms in sequential fashion in order of start times requested by the surgeons, by surgery specialty and duration estimates from surgeons, until the last surgery in the room can start before the end of the shift for the operating room.

Step 2: Assign one anesthesiologist to each room such that the anesthesiologist can perform most of the surgeries in the room.

Step 3: A few anesthesiologists are assigned to surgeries across rooms in order to ensure all surgeries have been assigned an anesthesiologist by specialty.

Step 4: If above plan cannot be implemented by anesthesiologists on regular duty, assign anesthesiologists from call.

While this practitioners heuristic is easy to understand and implement, it does not consider two important aspects. First, it does not explicitly consider uncertainty in surgical durations. Second, it does not directly use the feature that most anesthesiologists and operating rooms can perform more than one specialty. Thus, it does not exploit the flexibility in these resources. Define utilization as the fraction of the available shift time that is used by a particular resource. Inefficient assignment and scheduling of anesthesiologists and operating rooms to surgeries leads to low utilization and overtime of these resources. As seen in Figure 1, average daily utilization across the anesthesiologists is close to 0.75, with around 25% of days having an average daily utilization of less than 0.70. However, despite these lower levels of utilization, the average number anesthesiologists on call is around 6 per day. Similarly, for operating rooms the average daily utilization is close to 78% (Figure 2) but the average overtime per day is around 18 hours. Taken together average call and overtime costs for anesthesiologists and rooms at this department are about 33% of revenues. A more effective optimization based assignment and scheduling system that considers uncertainty in surgical durations and flexibility in the resources could potentially reduce overtime and on call costs.

<div align="center">**Insert Figure 1 and 2 about here**</div>

Assignment and scheduling decisions at this hospital are complicated by the large number of ORs and anesthesiologists, variety in surgical procedures, variability in anesthesiologist workload, and unpredictability in surgery durations. More details on these aspects are provided in the Electronic Companion. The management of the operating services department felt that the current planning process did not adequately consider these complicating factors. Thus, they believed that the daily expected resource usage and overtime costs across anesthesiologists and operating rooms could be considerably lowered by developing an optimization model, which led to our involvement. This model is formulated as a two-stage mixed-integer stochastic dynamic program with recourse. The first stage of this model allocates these resources across multiple surgeries with uncertain durations, and prescribes the sequence of surgeries to these resources. Assuming that each surgery should be scheduled as early as possible, this consequently provides a scheduled start time for surgeries. The second stage determines the actual start times to surgeries based on realized durations of preceding surgeries, and assigns overtime to resources to ensure all surgeries are completed using the allocation and sequence determined in the first stage. The size and complexity of the problem precluded solution using conventional methods. Therefore, we develop a data driven robust optimization approach that solves large-scale real-sized versions of this model close to optimality. Next, we describe the model formulation, present its properties and describe its solution techniques.

3. Model

We start by presenting a model formulation of the integrated anesthesiologist and room scheduling problem for surgeries. The planning horizon for this model is a single day. This was consistent with the requirements in the application context. Here, surgeries were finalized by the aggregate block scheduling system and released to the operating services department for detailed scheduling and staffing only the day before the surgery. Further, the availability of anesthesiologists on regular duty and on call was already determined by a longer range planning system at this department. Subsequently, the remaining decisions were the daily assignment of anesthesiologists and operating rooms to surgeries and determining the sequence of surgeries at these resources. The model makes these decisions in the previous day for the next day. The planning horizon of one day is also consistent with other research on operating room models (Cardoen et al. 2009a & 2009b and Batun et al. 2011). To provide a precise definition of the model, let $h, i, j \in I$ index the set of surgeries, $a \in A$ index the set of anesthesiologists, and $r \in R$ index the set of operating rooms. We define the following variables that are optimized:

x_{ia}: 1 if anesthesiologist a is assigned to surgery i, 0 otherwise

y_a: 1 if anesthesiologist a is assigned from call, 0 otherwise

z_{ir}: 1 if room r is assigned to surgery i, 0 otherwise

v_r: 1 if room r is assigned any surgery, 0 otherwise

u_{ij}: 1 if surgery i precedes surgery j, 0 otherwise

α_{ija}: 1 if surgery i and j are assigned to anesthesiologist a and i precedes j, 0 otherwise

β_{ijr}: 1 if surgery i and j are assigned to room r and i precedes j, 0 otherwise

s_i: Scheduled start time of surgery i (hrs)

S_i: Actual start time of surgery i (hrs)

$Over_a$: Overtime of anesthesiologist a (hrs)

$Over_r$: Overtime of room r (hrs)

In addition, let $\mathbf{x} = (x_{ia}) \ \forall i \in I, a \in A$, $\mathbf{y} = (y_a) \ \forall a \in A$, $\mathbf{z} = (z_{ir}) \ \forall i \in I, r \in R$, $\mathbf{u} = (u_{ij}) \ \forall i, j \in I$, $\mathbf{s} = (s_i) \ \forall i \in I$ denote the vectors associated with these variables. Next we define the following parameter or inputs:

κ_{ia}^A: 1 if anesthesiologist a can be assigned to surgery i, 0, otherwise

κ_{ir}^R: 1 if surgery i can be done in room r, 0 otherwise

g_a: 1 if anesthesiologist a is on regular duty, 0 otherwise

w_a: 1 if anesthesiologist a is on call, 0 otherwise

c_r: Fixed cost of opening operating room r (\$/day)

c_{oa}: Overtime cost of anesthesiologist a (\$/hr)

c_{or}: Overtime cost of room r (\$/hr)

c_q: Cost of assigning anesthesiologist from call (\$/day)

t_a^{start}: Start time of shift associated with anesthesiologist a (hr)

t_a^{end}: End time of shift associated with anesthesiologist a (hr)

T^{end}: End time of the day (hr)

M, $M_{seq}, M_{anesth}, M_{room}$: large positive numbers

The durations d_i of surgery i is uncertain $\forall i \in I$ and can be considered as a random variable. The vector of surgery durations for the day is represented by $\mathbf{d} = (d_i)$, $\forall i \in I$. We incorporate the uncertainty in surgery durations through a robust optimization approach, where we model d_i as an uncertain parameter that takes values in $[\bar{d}_i - \hat{d}_i, \bar{d}_i + \hat{d}_i]$. Here $\bar{d}_i$ is the nominal duration for surgery i and $\hat{d}_i$ is one sided maximum deviation for surgery i. We define the scaled deviations of d_i about its nominal value as $f_i = (d_i - \bar{d}_i)/\hat{d}_i$. Note that the scaled deviation f_i can take a value in $[-1, 1]$. Following the approach in Bertsimas and Sim (2004) and Denton et al. (2010) we subject the scaled deviations to a constraint $\sum_{i \in I} |f_i| \leq \tau$ so that, the total deviation across all surgeries is less than a known threshold τ. Here, τ bounds the total maximum deviation of surgery duration from the nominal value across all surgeries. This threshold is called the budget of uncertainty and represents the level of pessimism on the number of surgeries deviating from their nominal value. If $\tau = 0$, it is equivalent to solving the nominal value problem with $d_i = \bar{d}_i$, $\forall i \in I$.

The Integrated Anesthesiologist and Room Scheduling Problem (IARSP) consists of two stages. The first stage problem assigns anesthesiologists and rooms to surgeries, and prescribes a sequence of surgeries to be performed in each room and by each anesthesiologist. The second stage recourse function determines actual start times to surgeries based on realized durations of preceding surgeries, and assigns overtimes to resources such that all the surgeries are completed in the assignment and sequence prescribed by the first stage problem. Here, the two-stage approach assumes that all information about actual surgery durations is known early in the morning, which is of course not the case. However, this simplification has no impact on the solution since the only recourse action is to accumulate overtime without changing the sequences. Further, this simplification is consistent with the literature on surgery scheduling employing two-stage stochastic models with recourse (Denton et al. 2010, Batun et al. 2011, Mancilla and Storer 2012). The *IARSP* can be written as:

$$(IARSP) \qquad \mathcal{V}^*(\tau) = \min \left\{ \sum_{r \in R} c_r v_r + \sum_{a \in A} c_q y_a + \mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \right\} \tag{1}$$

subject to,

$$\sum_{a \in A} x_{ia} = 1 \qquad\qquad \forall i \in I \tag{2}$$

$$\sum_{r \in R} z_{ir} = 1 \qquad\qquad \forall i \in I \tag{3}$$

$$z_{ir} \leq v_r \qquad\qquad \forall i \in I, r \in R \tag{4}$$

$$x_{ia} \leq g_a + y_a \qquad\qquad \forall i \in I, a \in A \tag{5}$$

$$y_a \leq w_a \qquad \forall a \in A \qquad (6)$$

$$s_i \geq t_a^{start} - M(1 - x_{ia}) \qquad \forall i \in I, a \in A \qquad (7)$$

$$x_{ia} \leq \kappa_{ia}^A \qquad \forall i \in I, a \in A \qquad (8)$$

$$z_{ir} \leq \kappa_{ir}^R \qquad \forall i \in I, r \in R \qquad (9)$$

$$\alpha_{ija} \leq u_{ij} \qquad \forall i, j \in I, a \in A \qquad (10)$$

$$\beta_{ijr} \leq u_{ij} \qquad \forall i, j \in I, r \in R \qquad (11)$$

$$u_{ij} + u_{ji} \leq 1 \qquad \forall i, j \in I \qquad (12)$$

$$u_{ih} \geq u_{ij} + u_{jh} - 1 \qquad \forall i, j, h \in I \qquad (13)$$

$$\alpha_{ija} + \alpha_{jia} \leq x_{ia} \qquad \forall i, j \in I, a \in A \qquad (14)$$

$$\alpha_{ija} + \alpha_{jia} \leq x_{ja} \qquad \forall i, j \in I, a \in A \qquad (15)$$

$$\alpha_{ija} + \alpha_{jia} \geq x_{ia} + x_{ja} - 1 \qquad \forall i, j \in I, a \in A \qquad (16)$$

$$\beta_{ijr} + \beta_{jir} \leq z_{ir} \qquad \forall i, j \in I, r \in R \qquad (17)$$

$$\beta_{ijr} + \beta_{jir} \leq z_{jr} \qquad \forall i, j \in I, r \in R \qquad (18)$$

$$\beta_{ijr} + \beta_{jir} \geq z_{ir} + z_{jr} - 1 \qquad \forall i, j \in I, r \in R \qquad (19)$$

$$\alpha_{ija} \geq x_{ia} + x_{ja} + \beta_{ijr} - 2 \qquad \forall i, j \in I, r \in R, a \in A \qquad (20)$$

$$\beta_{ijr} \geq z_{ir} + z_{jr} + \alpha_{ija} - 2 \qquad \forall i, j \in I, r \in R, a \in A \qquad (21)$$

$$x_{ia}, y_a, z_{ir}, , u_{ij}, v_r, \alpha_{ija}, \beta_{ijr} \in \{0, 1\} \qquad \forall i, j \in I, r \in R, a \in A \qquad (22)$$

$$s_i \geq 0 \qquad \forall i \in I \qquad (23)$$

Objective function (1) consists of three terms. The first term is the fixed cost for opening operating rooms each day. The second term is the cost of assigning anesthesiologists from call. The third term $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$, represents the worst case second stage cost and is described in detail below. Constraints (2) and (3) assign each surgery exactly one anesthesiologist and one operating room respectively. Constraint (4) ensures that v_r is set to 1 whenever any surgery is assigned to operating room r. Constraint (5) ensures that an anesthesiologist can be assigned to a surgery only if they are on regular duty or on call. Constraint (6) enforces that an anesthesiologist can be assigned from call only if they are listed in the call list. Constraint (7) ensures that an anesthesiologist can be assigned a surgery only if the scheduled start time of the surgery is after the shift start time of the anesthesiologist. Constraints (8) and (9) ensure that surgeries are assigned rooms and anesthesiologists by specialty. Constraint (10) enforces the condition that if an anesthesiologist is used to conduct surgery i before surgery j, then surgery i has to precede surgery j or u_{ij} is set to 1. Constraint (11) imposes the similar condition and sets u_{ij} to 1 when surgery i precedes surgery j in an operating room Constraint (12) ensures that only one of u_{ij} or u_{ji} can be 1. Constraint (13) is required to maintain consistency of schedule between any three surgeries that follow each other, so that if i precedes j and j precedes h, then i should precede h. Constraints

(14)-(15) restrict that only one of α_{ija} and α_{jia} can be 1 only if both surgeries i and j are assigned to anesthesiologist a. Constraint (16) enforces either α_{ija} for α_{jia} is set to 1 if surgeries i and j are assigned to anesthesiologist a. In addition, constraint (16) ensures that the sequencing constraints for anesthesiologists α_{ija} is active only for those surgeries that are assigned to the same anesthesiologist. Constrains (17)-(19) are similar logical constraints corresponding to the sequencing of rooms. Constraints (20) and (21) maintain consistency of sequencing variables between operating rooms and anesthesiologists. Constraint (20) enforces that if anesthesiologist a and operating room r is assigned to surgeries i and j and i precedes j in operating room r then i has to precede j in assignment to anesthesiologist a. Constraint (21) is a similar constraint that makes sure that if surgery i precedes surgery j with anesthesiologist a, then i has to precede j in the assignment of operating room r. Constraints (22) and (23) represent variable domains.

The worst case second stage cost is given by:

$$\mathcal{Q}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s}) = \max_{\mathbf{d}\in\mathcal{D}(\tau)} \mathcal{R}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s},\mathbf{d}) \tag{24}$$

$$\mathcal{D}(\tau) = \left\{ \mathbf{d}\in\mathbb{R}^{|I|} : d_i = \bar{d}_i + f_i\hat{d}_i, i\in I, \mathbf{f}\in\mathcal{F}(\tau) \right\} \tag{25}$$

$$\mathcal{F}(\tau) = \left\{ \mathbf{f}\in\mathbb{R}^{|I|} : \sum_{i\in I}|f_i|\leq\tau, -1\leq f_i\leq 1 \right\} \tag{26}$$

$\mathcal{R}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s},\mathbf{d})$ is the total overtime cost across all resources, for a given assignment, sequence, scheduled surgery start times and surgery durations. This is maximized over the vector of surgery durations $\mathbf{d}$ to determine the worst case cost, where $\mathbf{d}$ is restricted to lie in the uncertainty set $\mathcal{D}(\tau)$ given by (25). This equation restricts d_i, the duration of surgery i, to lie within a maximum deviation of $\hat{d}_i$ from the nominal value of the duration $\bar{d}_i$. The total extent of such deviations is specified by the set $\mathcal{F}(\tau)$, which is defined by (26) and is well suited to our problem context. In particular, the effective allocation of multiple parallel resources such as anesthesiologists and rooms which are used repeatedly across the surgeries in a given day requires a specification of τ, an overall level or budget of uncertainty across surgical durations. This is enforced by (26), which specifies that the maximum deviation across all surgeries is at most τ. A schedule based on a large τ would be overly accommodating towards the second stage cost, while a schedule corresponding to a small τ would not be accommodating enough. In §4, we present a methodology to determine $\bar{d}_i$, $\hat{d}_i$ and τ based on historical data.

In determining $\mathcal{R}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s},\mathbf{d})$, it is important to note that the only decision variables at this stage are the actual start times of the surgeries and the overtime for the anesthesiologists and rooms. We pick these variables to minimize total overtime costs while ensuring that all the surgeries scheduled for the day are completed and there are no conflict in actual start times of surgeries assigned to the same resource. To compute $\mathcal{R}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s},\mathbf{d})$ we formulate the linear program:

$$\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}) = \min \left\{ \sum_{a \in A} c_{oa} Over_a + \sum_{r \in R} c_{or} Over_r \right\} \tag{27}$$

subject to,

$$S_j \geq S_i + d_i - M_{seq}(1 - u_{ij}) \qquad \forall i, j \in I \tag{28}$$

$$S_i \geq s_i \qquad \forall i \in I \tag{29}$$

$$Over_a \geq S_i + d_i - t_a^{end} - M_{anesth}(1 - x_{ia} + y_a) \qquad \forall i \in I, a \in A \tag{30}$$

$$Over_r \geq S_i + d_i - T^{end} - M_{room}(1 - z_{ir}) \qquad \forall i \in I, r \in R \tag{31}$$

$$S_i, Over_a, Over_r \geq 0 \qquad \forall i \in I, a \in A, r \in R \tag{32}$$

The objective function consists of the sum of overtime across all the resources. Constraint (28) ensures that the start time of the succeeding surgery is only after the end time of the preceding surgery. Constraint (29) ensures that the actual start time of the surgery can be no earlier than the scheduled start time. Constraints (30) and (31) define the overtime for anesthesiologists on regular duty and operating rooms respectively, which is the time difference between the end time of the last surgery in that shift and the regular shift end time for the resource. Constraint (32) restricts start time and overtime variables to be non-negative variables. The *IARSP* is a robust optimization model with the recourse represented by this linear program. We next develop some structural properties that is useful in constructing solution techniques for this model.

PROPOSITION 1. *The* IARSP *has relatively complete recourse.*

All proofs are provided in the Electronic Companion. Proposition 1 implies that for every feasible first stage solution, there exists a feasible second stage solution. This proposition allows us to evaluate second stage costs for every feasible first stage solution. This is important for the solution method for the *IARSP* described in §3.1.1. However, evaluating $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ for any given first stage solution requires one to solve the problem given in equations (24)-(32), which is not easy due to the max-min operator in its objective. Let λ_{ij}, ϕ_i, μ_{ia}, $\theta_{i,r}$ $\forall i, j \in I, a \in A, r \in R$ be dual variables corresponding to constraints (28)-(31) respectively. Further, define $\pi_i = \sum_{j \in I - \{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir}$, and $\xi_i = f_i \pi_i$. The following proposition simplifies the computation of $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ and consequently the *IARSP*.

PROPOSITION 2. *If parameter τ is chosen to be a positive integer, then $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ can be reformulated as the following mixed integer program.*

$$\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) = \max \left\{ \sum_{i \in I} \left(\bar{d}_i \pi_i + \xi_i \hat{d}_i \right) + \sum_{i \in I} s_i \phi_i - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij} (1 - u_{ij}) \right.$$

$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} (1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$

subject to,

$$\sum_{i \in I} \mu_{ia} \leq c_{oa} \qquad\qquad \forall a \in A \qquad (33)$$

$$\sum_{i \in I} \theta_{ir} \leq c_{or} \qquad\qquad \forall r \in R \qquad (34)$$

$$\sum_{\substack{j \in I \\ j \neq i}} \lambda_{ij} - \sum_{\substack{j \in I \\ j \neq i}} \lambda_{ji} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} - \phi_i \geq 0 \qquad\qquad \forall i \in I \qquad (35)$$

$$\sum_{j \in I - \{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} = \pi_i \qquad\qquad \forall i \in I \qquad (36)$$

$$\sum_{i \in I} f_i \leq \tau \qquad\qquad (37)$$

$$\xi_i \leq M_f f_i \qquad\qquad \forall i \in I \qquad (38)$$

$$\xi_i \leq \pi_i \qquad\qquad \forall i \in I \qquad (39)$$

$$\xi_i, \pi_i, \lambda_{ij}, \theta_{ir}, \mu_{ia}, \phi_i \geq 0 \qquad\qquad \forall i, j \in I, a \in A, r \in R \qquad (40)$$

$$f_i \in \{0, 1\} \qquad\qquad \forall i \in I \qquad (41)$$

As described in the Electronic Companion, the proof of this proposition follows from strong duality of the second stage recourse problem for a given first stage solution. Note that as a consequence of Proposition 2 in which τ is set to an integer, $f_i, \forall i \in I$ are also now binary variables. Thus, in the worst case, surgeries are set to either their nominal value or maximum positive deviation. In effect, τ can now be interpreted as the upper bound on the number of surgeries that reach their maximum deviation. Thus, restricting τ to be a positive integer as in this proposition allows for a more natural interpretation of τ, which was important in the application context. As discussed in §4, this interpretation drives our data driven method in setting a parametric value for τ from historical data. Propositions 1 and 2 imply that $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ can be evaluated for any given first stage feasible solution by solving a mixed integer program. In particular, let $(\boldsymbol{\pi}^l, \boldsymbol{\xi}^l, \boldsymbol{\lambda}^l, \boldsymbol{\phi}^l, \boldsymbol{\mu}^l, \boldsymbol{\theta}^l)$ be the solution to $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ for some given $(\mathbf{x}^l, \mathbf{y}^l, \mathbf{z}^l, \mathbf{u}^l, \mathbf{s}^l)$. Then, the following propositions provide a lower bound and characterize the structure of $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$. They will be used in the solution method provided in §3.1.1.

PROPOSITION 3. *A lower bound on* $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ *is provided by* $\sum_{i \in I} \left(\bar{d}_i \pi_i^l + \xi_i^l \hat{d}_i \right) +$ $\sum_{i \in I} s_i \phi_i^l - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij}^l (1 - u_{ij}) - M_{anesth} \sum_{i \in I a \in A} \mu_{ia}^l (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l (1 - z_{ir}) -$ $\sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l t_a^{end}$.

PROPOSITION 4. $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ *is a piecewise-linear convex function in the first stage decision variables* $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}$.

In light of Proposition 4, the *IARSP* now reduces to a piecewise-linear convex mixed integer program in which $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$, the convex part of the objective function can be evaluated by using Proposition 2

and solving a mixed integer program. However, given this non-linearity and the large number of integer variables in our application, the *IARSP* cannot be solved using powerful solvers for non-linear programs such as BARON (Sahinidis 2014) and DICOPT (Viswanathan and Grossmann 1990). Consequently, we develop the following model based heuristic procedure to solve this problem.

3.1. Solution Methods

We start by describing the model based heuristic. We then discuss sample average approximation based techniques that are commonly used in literature, which we use to benchmark the model based and practitioner's heuristics. The performance of these methods along with practitioner's heuristic (described in §2) will be discussed in §5.

3.1.1. Model Based Heuristic. This heuristic is based upon Kelley's algorithm (Kelley 1960) as described in Thiele et al. (2009) to solve robust optimization problems with recourse. Here, we consider the *IARSP* and in light of Proposition 4, approximate $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ by a piecewise linear equation via successive linear cuts. We then use this approximation in constructing the master problem at the kth iteration of the heuristic, $MP(k)$, defined as:

$MP(k)$

$$\min_{v_r, y_a, \phi} \left\{ \sum_{r \in R} c_r v_r + \sum_{a \in A} c_q y_a + \psi \right\} \tag{42}$$

subject to,

$$(2) - (23)$$

$$\psi \geq \sum_{i \in I} \left(\bar{d}_i \pi_i^l + \xi_i \hat{d}_i \right) + \sum_{i \in I} s_i \phi_i^l - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij}^l \left(1 - u_{ij} \right) - M_{anesth} \sum_{i \in I a \in A} \mu_{ia}^l \left(1 - x_{ia} + y_a \right)$$

$$- M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l \left(1 - z_{ir} \right) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l t_a^{end} \qquad l = 0, 1, 2..., k-1 \tag{43}$$

$$\psi \geq 0 \tag{44}$$

Observe that in this problem, we approximate the value of $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ by a variable $\psi \geq 0$. To improve this approximation, in each iteration of the heuristic we use (43) to enforce the condition that ψ is greater than or equal to the lower bound of $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$, as established by Proposition 3. This results in constraints (43) in which ψ approximates $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ by a piecewise linear equation via successive linear cuts. This problem will be used in the model based heuristic formalized by the following algorithm.

Model Based Heuristic

Step 1. Initialize $U \leftarrow \infty, L \leftarrow 0, k \leftarrow 0, l \leftarrow 0$. Set $\epsilon > 0$ to be sufficiently small.

Step 2. Solve the *MP(k)* and let the solution be $\tilde{x}_{ia}^k, \tilde{y}_a^k, \tilde{z}_{ir}^k, \tilde{u}_{ij}^k, \tilde{s}_i^k, \tilde{v}_r^k, \tilde{\alpha}_{ija}^k, \tilde{\beta}_{ijr}^k, \tilde{\psi}^k$. Set $L \leftarrow \sum_{r \in R} c_r \tilde{v}_r^k +$
$\sum_{a \in A} c_q \tilde{y}_a^k + \tilde{\psi}^k$.

Step 3. Compute $\mathcal{Q}(\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\mathbf{z}}^k, \tilde{\mathbf{u}}^k, \tilde{\mathbf{s}}^k)$ for given $\tilde{x}_{ia}^k, \tilde{v}_r^k, \tilde{z}_{ir}^k, \tilde{u}_{ij}^k, \tilde{s}_i^k$, obtained in **Step 2** by solving the mixed integer programming formulation given in Proposition 2. Let the optimal solution be $\xi_i^k, \pi_i^k, \lambda_{ij}^k, \theta_{ir}^k, \mu_{ia}^k, \phi_i^k$. $U \leftarrow \min \left\{ U, \sum_{r \in R} c_r \tilde{v}_r^k + \sum_{a \in A} c_q \tilde{y}_a^k + \mathcal{Q}(\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\mathbf{z}}^k, \tilde{\mathbf{u}}^k, \tilde{\mathbf{s}}^k) \right\}$.

Step 4. If $U - L < \epsilon$ go to **Step 6**, **else** go to **Step 5**.

Step 5. $k \leftarrow k + 1$. Add constraint $\psi \geq \sum_{i \in I} \left(\bar{d}_i \pi_i^k + \xi_i \hat{d}_i \right) + \sum_{i \in I} s_i \phi_i^k - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij}^k (1 - u_{ij})$
$- M_{anesth} \sum_{i \in I a \in A} \mu_{ia}^k (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^k (1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^k T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^k t_a^{end}$ to the *MP(k)*.
Go to **Step 2**.

Step 6. $\tilde{x}_{ia}^k, \tilde{y}_a^k, \tilde{z}_{ir}^k, \tilde{u}_{ij}^k, \tilde{v}_r^k, \tilde{s}_i^k, \tilde{\alpha}_{ija}^k, \tilde{\beta}_{ijr}^k$ is the heuristic solution to *IARSP*.

The above algorithm is well suited for the IARSP as from Proposition 4, its objective is piecewise-linear convex and cutting plane methods such as those used in Step 5 of this algorithm are finitely convergent for piecewise linear functions (Ruszczyński 2006). Also, in this algorithm, in early iterations, the solution in Step 2 is obtained by employing the *user callbacks* feature of the solver used to solve *MP(k)*. Here, instead of solving this problem to optimality in the initial iterations, we request the solver to return a feasible solution, which is then used to apply cuts in Step 5 and approximate the convex function $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ at each corresponding feasible solution. We do so because *MP(k)* is a problem with a large number of integer variables and solving it to optimality can be computationally expensive with poor returns at early iterations when $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ has not been approximated well enough by constraints (43). Results on the computational performance and the time required for this heuristic is provided in Section §5.

3.1.2. Benchmark Heuristic. Here we use Sample Average Approximation (SAA) based methods similar to those provided in Denton et al. (2007) to benchmark the model based heuristic. In SAA, instead of using the worst case formulation, we solve with expected second stage costs. The expectation is based on scenarios drawn from an estimated distribution of surgeries. The resultant two-stage stochastic optimization problem is solved by the *L-Shaped* method (Birge and Louveaux 1988). We describe the model formulation of the SAA version and provide details on the estimation of the distribution in the Electronic Companion. However, since the sample average based method was unable to solve large-scale problems such as those found in the application, we use this method on smaller problems constructed from real data. We provide the results of the comparison between the SAA based method and the robust optimization based method in §5.

4. Parameter Estimation and Model Calibration

In this section we use historical data of surgery durations to choose the uncertainty sets $\mathcal{D}(\tau)$ and $\mathcal{F}(\tau)$. The performance of robust optimization depends closely on the definition of these uncertainty sets. If the optimal values of d_i, $\forall i \in I$ in the inner maximization problem $(\max_{\mathbf{d} \in \mathcal{D}(\tau)} \mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}))$ are significantly larger than the corresponding $\bar{d}_i$, the resulting first stage problem will be overly pessimistic towards the realization of surgery durations. This may lead to higher first stage costs. Conversely, if the optimal values of d_i, $\forall i \in I$ are too close to the corresponding $\bar{d}_i$, the uncertainty sets would not cover many cases of future realizations in which the surgery durations deviate significantly from the nominal value. This could result in higher second stage costs. Thus, we need to look at the combined first and second stage costs while designing the uncertainty sets. Designing the uncertainty sets involves setting the following parameters: the nominal surgery duration $\bar{d}_i$, $\forall i \in I$, the maximum deviation $\hat{d}_i$, $\forall i \in I$, and the robust optimization parameter τ.

There have been several approaches suggested for designing uncertainty sets. Ben-Tal et al. (2009) provide the theoretical background for deciding good uncertainty sets. Denton et al. (2010) use the 10th and 90th percentile width of historical surgery durations as the width $[\bar{d}_i - \hat{d}_i, \bar{d}_i + \hat{d}_i]$ in an operating room assignment application. Subsequently, they perform sensitivity analysis and calibrate their model to an equivalent SAA based solution to decide the robust optimization parameter τ. Bertsimas et al. (2013) propose using statistical hypothesis tests to construct uncertainty sets. Denton et al. (2010) and Bertsimas et al. (2013) model the uncertainty sets based on historically observed values of a *single* uncertain parameter. In our application with a wide variety of surgery specialties with considerable variability in surgery durations across specialties, a percentile width not conditional on surgery characteristics would be unnecessarily wide, leading to an overly pessimistic uncertainty set. Therefore, we incorporate these characteristics and propose a joint estimation and calibration procedure to design the uncertainty set. Our procedure provides tight uncertainty sets that take into account observable surgery characteristics while making no assumptions on the probability distribution of surgery durations. We further calibrate the uncertainty set by evaluating the performance of the robust solution to empirical realizations.

There were two data sets available to us. The first data set $\Delta^E = \left\{ \widetilde{d}^{(m)}, \widetilde{\mathbf{b}}^{(m)} \right\}_{m=1}^{M}$ consists of $M = 25700$ samples of $\widetilde{d}^{(m)}$ corresponding to the historical realization of durations of surgery m and $\widetilde{\mathbf{b}}^{(m)}$ which represents the observed characteristics of surgery m. Table 2 provides details on the surgery characteristics included in $\widetilde{\mathbf{b}}(m)$, $\forall m$ and the variable names used for the subsequent regression. The second data set Δ^C was partitioned into disjunctive training and testing sets, $\Delta^{C-Train} = \left\{ \widetilde{\mathbf{d}}^{(n)}, \widetilde{\mathbf{b}}^{(n)} \right\}_{n=1}^{N_1}$ and $\Delta^{C-Testing} = \left\{ \widetilde{\mathbf{d}}^{(n)}, \widetilde{\mathbf{b}}^{(n)} \right\}_{n=1}^{N_2}$. For these data sets, $N_1 = 120$ days and $N_2 = 60$ days. Both these data sets consist of $\widetilde{\mathbf{d}}^{(n)}$ representing the vector of realized durations and $\widetilde{\mathbf{b}}^{(n)}$ denoting the vector of surgery characteristics of all surgeries performed on day n.

Insert Table 2 about here

We use Δ^E and $\Delta^{C-Train}$ to estimate $\boldsymbol{\mathcal{D}}(\tau)$ and $\boldsymbol{\mathcal{F}}(\tau)$. Most of the research in estimating uncertainty sets is for single stage problems when feasibility is not guaranteed. Since we have two stages and second stage feasibility is guaranteed by Proposition 1, we develop the following procedure that comprises of an estimation and a calibration step.

Step 1: Estimation

First, for a given parameter $\rho \in (0,1)$, we define conditional quantile functions $g_L(\mathbf{b}; \rho)$ and $g_U(\mathbf{b}; \rho)$ such that:

$$\mathbf{P}\left[\widetilde{d} \le g_L(\widetilde{\mathbf{b}}; \rho)\right] = \frac{1-\rho}{2}, \text{ and } \mathbf{P}\left[\widetilde{d} \ge g_U(\widetilde{\mathbf{b}}; \rho)\right] = \frac{1-\rho}{2}$$

Thus, given observed surgery characteristics $\widetilde{\mathbf{b}}$ the future realization $\widetilde{d}$ will lie in the set $\left[g_L(\widetilde{\mathbf{b}}; \rho), g_U(\widetilde{\mathbf{b}}; \rho)\right]$ with probability ρ. The true quantile functions are not known to us, so we obtain estimates of the quantile functions $\hat{g}_L(\widetilde{\mathbf{b}}; \rho)$ & $\hat{g}_U(\widetilde{\mathbf{b}}; \rho)$ through a conditional quantile regression method (Koenker 2005) applied on the data set Δ^E. The use of conditional quantile regression for estimating uncertainty sets has been recently proposed by Tulabandhula and Rudin (2014). Quantile regression estimates the quantiles of the response variable (i.e., the surgery durations), given certain values of the predictor variables. Quantile regression has several advantages over the commonly used ordinary least squares (OLS) regression. First, this approach suits our application better since our objective is to find upper and lower bounds on surgery durations such that future realization would lie within this bound with a given probability. Conditional quantiles provide these bounds without making any assumption on the probability distribution of surgery durations. Second, quantile regression is more robust to outliers and third, it does not assume the dispersion of the response variable to be independent of the predictor variables.

We perform quantile regression using the `quantreg` package available in `R` (Koenker 2013). The response variable is the realized surgery duration. The possible set of predictors are the surgeon's estimate of surgery duration, the fitness level of the patient prior to the surgery measured by the American Society of Anesthesiologist (ASA) score[1], the age of the patient, whether the patient is an inpatient or outpatient, the specialty of the surgery, the surgeons name, the services provided indicated by the type of the Current Procedure Terminology (CPT)[2] codes used and the number of CPT codes used by the surgeons. For the CPT codes and surgeon's names we cluster the variables via a k-means clustering similar to He et al. (2012). This clustering was done in order to account for the large number of factors in these variables and to avoid over-specification of the model. The details of the clustering procedure followed is provided in the Electronic Companion.

---

[1] https://www.asahq.org/resources/clinical-information/asa-physical-status-classification-system

[2] http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page

The selection of variables was done comparing the Akaike Information Crietria (AIC) and the Mean Square Prediction Error (MSPE). The results of these tests are provided in the Electronic Companion. We also checked for collinearity using Variance Inflation Factors (VIF) following the criteria discussed in (Hair et al. 2006, pp. 191-193). Highly collinear variables (i.e. with VIF ≥ 10) were removed. For example, we found that the specialty of surgery had a VIF of 40.1 as it was collinear with CPT codes, as these codes were specific to a specialty. On performing these tests we found that the ASA score, the surgeon's estimate of duration, patient class (Inpatient, Outpatient or Same Day Admit), clustered variables corresponding to surgeons and CPT codes were significant. The ASA score is a strong indicator of increasing complexity of the surgical procedure since it is an indicator of the level of fitness of the patient before coming into the surgery. A patient with an ASA score of 1, implying the patient is a healthy person would be expected to demonstrate less complications during surgery, while a patient with an ASA score of 3 (with severe systemic disease) would be expected to have more complications during surgery. We found that the coefficient of ASA score was -0.023 at the 0.1 quantile and 0.122 at the 0.9 quantile. Thus, every increase in ASA score contributes to approximately 7.3 minutes ($\approx 0.122\,\text{hrs}$) of additional surgery time at the 0.9 quantile level, while the effect of increment in ASA score is negligible for very short surgeries. This is intuitive as the negative effects of patient fitness would be significant for longer surgeries and would not be as impactful for shorter surgeries. As expected, the surgeon's estimate of surgery duration would be strongly correlated with the actual duration and would explain variance not captured by other variables, since there are several factors that the surgeon is aware of that are not captured by other available data. However, as explained previously there is some error in surgeon's estimates as well. We found that surgeon's estimate was on an average 12 minutes higher than actual surgery durations. Also the coefficient of surgeons' estimates in the quantile regression model was smaller for shorter surgeries than for longer surgeries. This is because surgeons tend to be more accurate in their estimates for longer surgeries than for shorter surgeries. One possible explanation is that it was observed that surgeons tend to round to the nearest quarter of an hour while providing their estimates. This leads to an error which is more pronounced for shorter surgeries than for longer surgeries. We also found that clusters of surgeons are significant because as described in the Electronic Companion these in effect represented the experience level of surgeons. Finally, as anticipated, the type of surgery itself with its associated CPT code affected surgical durations. However, the number of CPT codes was not significant as they could be associated with relatively simpler sub-procedures common to all surgeries. Similarly, the age of the patient was not significant as it was captured in the surgeons estimate of duration.

Once we have obtained the estimated conditional quantile functions, for each surgery $i \in I$ we set, $\bar{d}_i + \hat{d}_i = \hat{g}_U(\widetilde{\mathbf{b}}_i; \rho)$, $\bar{d}_i - \hat{d}_i = \hat{g}_L(\widetilde{\mathbf{b}}_i; \rho)$. This gives,

$$\bar{d}_i = \frac{\hat{g}_U(\widetilde{\mathbf{b}}_i; \rho) + \hat{g}_L(\widetilde{\mathbf{b}}_i; \rho)}{2} \text{ and } \hat{d}_i = \frac{\hat{g}_U(\widetilde{\mathbf{b}}_i; \rho) - \hat{g}_L(\widetilde{\mathbf{b}}_i; \rho)}{2}$$

Define $\tau' \in [0,1]$ so that $\tau = \lfloor \tau'|I| \rfloor$. Here, τ' represents the fraction of total surgeries in a given day $|I|$ which have reach their maximum duration. We then substitute the above equations in (25) and (26) for observed surgery characteristics vector $\mathbf{b}_i$ and *given* parameters $\rho \in [0,1]$ and $\tau' \in (0,1)$. Then the uncertainty sets are given by:

$$\boldsymbol{\mathcal{D}}(\tau) = \boldsymbol{\mathcal{D}}(\rho,\tau') = \left\{ \mathbf{d} \in \mathbb{R}^{|I|} : d_i = \frac{\hat{g}_U(\widetilde{\mathbf{b}}_i;\rho) + \hat{g}_L(\widetilde{\mathbf{b}}_i;\rho)}{2} + f_i \frac{\hat{g}_U(\widetilde{\mathbf{b}}_i;\rho) - \hat{g}_L(\widetilde{\mathbf{b}}_i;\rho)}{2}, i \in I, \mathbf{f} \in \boldsymbol{\mathcal{F}}(\tau') \right\} \quad (45)$$

$$\boldsymbol{\mathcal{F}}(\tau) = \boldsymbol{\mathcal{F}}(\tau') = \left\{ \mathbf{f} \in \mathbb{R}^{|I|} : \sum_{i \in I} |f_i| \leq \lfloor \tau'|I| \rfloor, -1 \leq f_i \leq 1 \right\} \quad (46)$$

Step 2: Calibration

If we had full information on the surgery durations (i.e., if they were observable *ex ante*), and a deterministic solution could be executed, the resulting cost obtained when there is full information would be a lower bound to any heuristic solution. In stochastic programming this is referred to as the *wait-and-see* solution. The full information cost on day n is given as:

$$\mathcal{W}^{FI(n)} = \min \left\{ \sum_{r \in R} c_r v_r + \sum_{a \in A} c_q y_a + \sum_{a \in A} c_{oa} Over_a + \sum_{r \in R} c_{or} Over_r \right\} \quad (47)$$

subject to,

$$(2) - (23)$$

$$(28) - (32)$$

We solve $\mathcal{W}^{FI(n)}$ for each day in $\Delta^{C-Train}$ with $\mathbf{d} = \widetilde{\mathbf{d}}^{(n)}$.

The first stage variables for day n obtained by solving *IARSP* for day n using the model based heuristic are $(\mathbf{x}^{*(n)}, \mathbf{y}^{*(n)}, \mathbf{z}^{*(n)}, \mathbf{u}^{*(n)}, \mathbf{s}^{*(n)})$. The cost of the model based heuristic under a realized duration vector $\widetilde{\mathbf{d}}^{(n)}$ is defined as:

$$\mathcal{W}(\mathbf{x}^{*(n)}, \mathbf{y}^{*(n)}, \mathbf{z}^{*(n)}, \mathbf{u}^{*(n)}, \mathbf{s}^{*(n)}; \rho, \tau', \widetilde{\mathbf{d}}^{(n)}) = \sum_{r \in R} c_r v_r + \sum_{a \in A} c_q y_a + \mathcal{R}\left(\mathbf{x}^{*(n)}, \mathbf{y}^{*(n)}, \mathbf{z}^{*(n)}, \mathbf{u}^{*(n)}, \mathbf{s}^{*(n)}, \widetilde{\mathbf{d}}^{(n)}\right)$$

This represents the cost that would be realized at the end of day n if the model based heuristic was implemented with uncertainty set $\boldsymbol{\mathcal{D}}(\rho,\tau')$. The average performance of the model based heuristic across N samples relative to the full information case is defined as:

$$\bar{\mathcal{W}}(\rho,\tau') = \frac{1}{N_1} \sum_{n=1}^{N_1} \frac{\left[\mathcal{W}(\mathbf{x}^{*(n)}, \mathbf{y}^{*(n)}, \mathbf{z}^{*(n)}, \mathbf{u}^{*(n)}, \mathbf{s}^{*(n)}; \rho, \tau', \widetilde{\mathbf{d}}^{(n)}) - \mathcal{W}^{FI(n)}\right]}{\mathcal{W}^{FI(n)}} \quad (48)$$

We calculate $\bar{\mathcal{W}}(\rho,\tau')$ for several values of $\rho \in (0,1)$ and $\tau' \in [0,1]$ and choose the pair that minimizes $\bar{\mathcal{W}}(\rho,\tau)$. This is summarized in Table 3.

Insert Table 3 about here

From Table 3 we can see $\rho = 0.95$ and $\tau' = 0.2$ is optimal. This implies that at a 95% confidence level, we can set 20% of all surgeries to its maximum durations on any given day when we define the uncertainty sets $\mathcal{D}(\tau)$ and $\mathcal{F}(\tau)$ and solve the *IARSP*. At this value the model based heuristic solution was 24% more than the full information solution.

We used these values of ρ and τ' to evaluate the performance of model based heuristic relative to full information case as defined in (48) for the testing data set $\Delta^{C-Testing}$. Here, we found that the model based heuristic was 28% more than the full information solution. Thus, the out of sample performance, using $\Delta^{C-Testing}$ was close to the in sample performance using $\Delta^{C-Train}$. This provides validation to use these values of ρ and τ' in the computational analysis described next.

5. Computational Analysis

In this section, we conduct a computational analysis to evaluate our approach. To perform this analysis, we used data provided by the UCLA RRMC on all surgeries conducted in their operating room suite over a 14 month period. This analysis was essential to provide confidence in our method. Our computational analysis is divided into two sections. In §5.1 we evaluate the performance of the practitioner's heuristic and the procedures described in §3.1. In §5.2 we compare the performance of the model based heuristic with the actual resource assignment and scheduling decisions made at this hospital and estimate the cost savings.

5.1. Performance Evaluation

The size and scope of the scheduling activities during this time period demonstrated considerable variation as shown in Table 4. To ensure that our computational analysis captured this range of variation, we constructed five problems of varying sizes as shown in Table 5. For each of these five sets of problem instances we considered different values for c_q, c_{or} and c_{oa}. The actual value of c_q, c_{oa} and c_{or} at UCLA RRMC were $1000 per day, $150 per hour, and $450 per hour respectively. In addition to these actual values we considered values where we scaled one of these costs by a factor of 2 or 1/2 while keeping the other two at the current value. This led to 7 possible combinations of costs for each of the five problem instances and a total of $7 \times 5 = 35$ possible problems. We tried to solve the IARSP for these data sets using leading commercial solver for stochastic programs such as `ddsip` (Märkert and Gollmer 2008). However, other than the smaller problem instances A and B, these solvers could not even generate feasible solutions after more than 24 hours of computation, and the runs were aborted. This provides validation for developing the model based heuristic to solve the *IARSP*.

<div align="center">**Insert Table 4 and Table 5 about here**</div>

The heuristic procedures were coded in Python programming language (van Rossum 2001). The computational analyses were run on a workstation with 3.8 GHz AMD A10 processor, 8 GB of RAM, and Linux Mint as the operating system. For the MIP subroutine calls we used Gurobi 5.63 (Gurobi Optimization 2015) called from Python via the Gurobi Python Interface. In all the computations using the

model based heuristic, we set the gap $\epsilon = 5\%$. Thus, all the solutions of the model based heuristic were within a 5% gap from the lower bound and were solved within 25 minutes.

Tables 6 and 7 summarize the results obtained for the computational analysis. In Table 6 the performance of the model based heuristic and the practitioner's heuristic procedure is compared with the cost of the SAA based solution for small-scale problems. This table shows that these procedures are all very close to the SAA method and this does not change with changes in the cost parameters. In Table 7 we consider the more realistic medium and large-scale problems. Since SAA is unable to solve these problems, we provide the performance of the model based heuristic with respect to the practitioner's heuristic. From Table 7, we note that for these problems the model based heuristic provides significant cost reductions over the practitioner's heuristic. In particular, the percentage cost reduction for these problems ranged from 2.26% to 7.56% averaging around 4.95%.

<p style="text-align:center">**Insert Table 6 and Table 7 about here**</p>

We can also observe from Table 7 that the gains of the model based heuristic over the practitioner's heuristic improves as the size of the problem increases. This is because for smaller sized problems there are limited options and it is more likely the practitioner's heuristic achieves a solution that is close to optimal. Further, since in small sized problems $|I|$ is low, the number of surgeries that reach its worst case duration (i.e. $\tau = \tau'|I|$) is also low. In these circumstances, the solution of the model based heuristic and the practitioner's heuristic are similar and close to the nominal value solution, where surgical durations are set to its nominal duration estimates. However, as the problem size increases, the number of surgeries reaching its worst case duration increases. Under these circumstances, the practitioner's heuristic is outperformed by the model based heuristic, as the optimization inherent to the model based heuristic is more effective in utilizing resources that can be shared across multiple specialties and procedures. Finally, note that increasing the on call costs leads to the practitioner's heuristic doing much worse as this heuristic opts for increasing the number of on call anesthesiologists rather than trading off on call costs against overtime costs.

5.2. Model Validation

The objective of model validation was the demonstrate that the model based heuristic provides tangible cost savings over current practice. This was an essential step in convincing the management of the Operating Services department to implement our method. We performed model validation in two stages. In the first stage the cost savings were computed using historical data. In the second stage, we conducted live validation, where we compared in real time our decisions with those made at this hospital. Note that while conducting these validations, the heuristic had precisely the same information the planners at the UCLA RRMC had at the point of planning.

In the historical validation, we took 80 sample days, such that we covered the range of problem sizes encountered. These 80 samples were divided into 5 sets as described in Table 8. We next calculated the

average costs obtained by the model based heuristic and the costs resulting from the actual assignment and sequencing that was done by the RRMC planners. This reduction in costs across the five problems are also reported in Table 8 and this shows that the benefits of using the model based heuristic was significant and was increasing in problem size.

<div align="center">**Insert Table 8 and Table 9 about here**</div>

The real-time live validation was conducted over a four-week period. The number of surgeries per day over this period was similar to the range of problem sizes observed historically as is shown in Table 8. The results for the live validation is given in Table 9. This table shows that our heuristic reduced costs from current practice on average from 6.4% to 9.1% on 16 out of the 28 days corresponding to weekdays which were not holidays. This implied an estimated annual cost savings between \$2 million and \$2.86 million. It is also important to note that the practitioner's heuristic and the model based heuristic provided the same solution in the weekends, where the number of surgeries conducted are low, and both these methods provided solutions corresponding to the nominal value solution.

The model based heuristic outperforms the current practice in both historical and live validation for the following reasons. First, the nominal values obtained via quantile regression procedure provided better predictors for the realized surgical duration than the surgeon's estimates. Second, on average, around 50% of the surgeries exceeded the nominal value. This required an increase in realized workload from the nominal workload. We found that this increase can be effectively achieved by setting $\tau' = 20\%$. This led to the model based heuristic operating with fewer operating rooms and fewer anesthesiologists than actually used at the hospital, since these resource assignments were based on trading off the fixed costs for these resources with the chance of incurring overtime. Third, on average in 60% of the surgeries, the surgeon's estimate of surgery duration exceeded the realized duration. The planners chose additional resources and avoided overtime based on these quoted times. Thus, the associated plans tended to incur more resource usage costs (comprising of the fixed cost of opening an operating room and anesthesiologist call cost) rather than overtime costs in comparison to the model based heuristic. However, since this decision to use more resources was made without explicit consideration of overtime costs and errors in the duration estimate provided by the surgeons, this often led to greater total costs. In sum, the model based heuristic outperforms current practice due to better prediction and a more effective scheduling policy. The proportion of the gains due to each of these aspects are analyzed and summarized in the Electronic Companion.

Generally, in a stochastic decision problem, it is not valid to judge the quality of a decision based on an outcome, as due to randomness a good outcome does not necessarily imply a good decision. However, in this work, since the evaluation and validation of the model based heuristic have been extensive, we were confident that they would perform well in the real application. In the final analysis, the real measure of performance of this heuristic is the quality of the decision based on its solution, a question we consider next in the application.

6. Application

6.1. Implementation

We have implemented the model based heuristic as a decision support system at the operating services department of the UCLA Ronald Reagan Medical Center. Details of this system are provided in the Electronic Companion. The results before and after the implementation across key operational metrics and costs are summarized in Table 10. This table shows after implementation, the average number of anesthesiologists on call decreased by 6.7%, and average overtime hours for the anesthesiologists on regular duty reduced by 3.7%. This contributed to an increase of average daily utilization across the anesthesiologists by 3.5%. Similarly, the average number of operating rooms used decreased by 8.6%, and the average overtime hours at the operating rooms was reduced by 2.7%. This led to an increased average daily utilization across the operating rooms by 3.8%. The improvements in these operational metrics reduced average daily operating room costs by 8.6%, average daily overtime costs by 2.7%, and average daily call costs by 8.5%. This translates to an overall average daily cost saving of 7% or estimated to be $2.2 million on an annual basis.

The model based heuristic improved upon decision making at operating services due to two main reasons. First, it was more effective at utilizing the flexibility in the resources. Most anesthesiologists and operating rooms can perform more than one specialty, typically a primary and a secondary specialty. The model identified these operating room/anesthesiologist combinations and allocated surgeries across these different specialties to them. This led to better usage of resources than the previous approach, in which, surgeries from a single specialty were assigned to an operating room and anesthesiologist as much as possible. A surgery of a different specialty was assigned to an operating room only when there were high volume of surgeries in a particular day and this was often done without explicit consideration of the allotted anesthesiologists' specialty. Thus, this often required a separate anesthesiologist to perform these surgeries, who were often assigned from call and this was costlier. Second, the model based heuristic explicitly considered uncertainty in surgical durations while determining the daily schedule of an operating room. By using the estimation module, it determined which surgeries could be longer and more uncertain, and which surgeries could be shorter and more certain. It then combined long uncertain surgeries with short certain surgeries to effectively utilize gaps in the schedule in each operating room. This in turn reduced the number of operating rooms each day with the resulting cost reduction being more than any potential increases in overtime costs, thus reducing total costs. In contrast, the previous approach used surgeons' predictions of surgery durations. To compensate for the errors in these predictions, planners often underutilized operating rooms by leaving sufficient gaps between surgeries. This was done as they did not want to create delays from scheduled start times of succeeding surgeries and incur overtime costs. However, this often led to a larger number of operating rooms being used each day and consequently higher total costs.

Finally, we considered the impact of the schedules generated by our approach on the surgeons. While surgeons were not part of the operating services department, they are a critical element in the system. First, we computed the average idle time between surgeries and found that it reduced by 8 minutes after our work. The surgeons did not find this reduction significant enough to be disruptive, and in fact some of them preferred this as it made their schedule more efficient. Second, we calculated the average number of surgeons per OR per day. Prior to our work, on an average there were 1.54 surgeons per OR per day. After implementing the decision support system, there were 1.57 surgeons per OR per day. This marginal increase suggests that most of the benefits of our approach come from making the correct assignment of operating rooms and anesthesiologist to surgeries, and not from increasing the number of surgeons per OR per day. Both these aspects were important to verify that the surgeons were not inconvenienced by the model based approach.

6.2. Managerial Insights

We used the model based heuristic to generate several managerial insights. First, we considered the impact of reducing variability in surgical durations. In practice, this could be achieved by better procedures such as check lists, improved information technology, following the correct sequence in tasks and standardized operating protocols derived from best practices. These measures have been advocated by surgeons (Bates and Gawande 2003, Haynes et al. 2009, Gawande 2010). In addition, variance can be reduced by improving the prediction of surgical durations. This would require dividing the surgical process into a series of steps (such as time to incision, skin to skin and closure to exit), and predicting each segment individually as different patient characteristics affect each segment differently (Hosseini et al. 2014). The accuracy of this prediction can be improved by collecting more data on patient characteristics and surgeon experience (Kougias et al. 2012). To consider the impact of variance reduction, we started with the current level of standard deviation in surgical durations, and systematically reduced the standard deviation of the distribution of surgical durations across all surgeries by a fixed value. We used these modified distributions to simulate realizations of surgical durations. We then used these data sets to solve the IARSP using the model based heuristic and calculated the resulting total resource usage and overtime costs. These results are summarized by Figure 3. This figure shows that the benefits of further reduction in variability decreases, and that there are significant diminishing returns on reduction of variability. This suggests that rather than invest in capital-intensive medical equipment to achieve radical reductions in variability in surgical durations, the major cost benefits can be gained by focusing on incremental reduction in variability. This can be potentially achieved by better procedures and more detailed data collection for improved predictive analytics.

<div align="center">**Insert Figure 3 about here**</div>

Second, we consider the impact of allowing surgeries to start in operating rooms after 3 pm but before the end of the late shift of the anesthesiologists at 7 pm. This would require additional fixed technician

and nurse staffing costs. Such extensions can be considered if surgical demand on any day is significantly larger than average daily surgical demand. To perform this analysis, we considered four levels of demand corresponding to increases from daily average demand in surgeries that could occur during the days of any given week. For these scenarios, we incrementally increased the number of operating rooms available after 3 pm by one unit. We then calculated the resulting change in total resource usage and overtime costs from the case when we do not start surgeries after 3 pm but only use the day shift with additional rooms to accommodate such increases in demand. These results summarized in Figure 4 suggest that it is beneficial to allow such extensions and the number of operating rooms used depends on the level of demand. This analysis helps management understand how best to react to different levels of daily surgical demand and also estimate the corresponding changes in total costs.

Insert Figure 4 about here

Finally, we examined the benefit of increasing cross functionality of the operating rooms. To do so, we considered the various specialties and calculated the potential reduction in costs if the number of operating rooms available for each specialty described in Table 1 is increased. In practice, such increases can be achieved by investing in special equipment to convert general surgery operating rooms to have the cross functionality to accommodate a particular specialty. These results described in Figure 5 show that as we increase the number of operating rooms that could be used for a particular specialty, this can lead to a significant reduction in total resource usage and overtime costs, and these benefits are often more pronounced in certain specialties. This analysis forms a basis to identify such specialties, and determine the priority in which these operating rooms should be made cross functional to enable these additional rooms for the specialties. Further, an additional advantage of making operating rooms cross functional were that a higher number of daily surgeries could be more effectively accommodated without conducting new surgeries after 3 pm. In particular, we found this approach led to at least an additional 5% reduction from the lowest costs attainable for all the demand scenarios considered in Figure 4. This provides further justification for management to make the operating rooms more cross functional.

Insert Figure 5 about here

6.3. Qualitative Impact

The organizational impact of our work has been significant. Prior to our work, simple rules were used to make important decisions on allocation of anesthesiologists and rooms to surgeries and determining surgery start times. These rules developed based on experience and anecdotal evidence worked well during holidays and weekends when the number of surgeries conducted were low. However, as shown in Section 5 and observed during the implementation, our model significantly outperformed current practice during other days where the number of surgeries performed was high, and this resulted in considerable cost savings. Thus, our work demonstrated the value of model based approach and operations research

methods in dealing with complexity. This has encouraged the management to investigate other problems in this department using a structured and rigorous approach by employing operations research-based methodologies.

The managerial insights generated from our model have also contributed to the organizational impact. While the effect of variance reduction on improved clinical outcomes has been extensively documented (Neuhauser et al. 2011), our analysis showed that this could also reduce costs. This provided management with the further impetus to implement six sigma programs (Cima et al. 2011) to reduce variability at this department. In addition, our analysis provides management with clear guidance on when to start new surgeries after the day shift and in how many rooms. This provides them with a practical approach to mitigate the impact of varying levels of daily surgical demand on costs and is currently under consideration for implementation in the short term. Finally, we showed the benefits of making some operational rooms cross functional and how to prioritize implementation among the specialties. Furthermore, we demonstrate that this could potentially be a very effective way to accommodate changes in daily surgical demand. While the management at the operating services department was intrigued by this analysis, they felt that there could be significant investments required and this could also lead to disruptions in the schedule while some operating rooms were being reconfigured. Therefore, they are considering this initiative as part of the next broader hospital renovation project.

6.4. Limitations

This work has the following limitations. First, the estimation of uncertainty sets can be improved with additional data on the duration of each step in a surgery. However, this data was not available in our application. Second, we do not explicitly consider requests from surgeons for particular start times on a given day and for specific anesthesiologists. While these aspects can be easily incorporated in our model, the management felt that accommodating these requests explicitly can make the overall schedule inefficient and could create additional costs. Therefore, they preferred to make changes to the output in the decision support system only in the most exceptional circumstances. Third, we assume that the overtime payment sufficiently compensates staff for extended shifts. However, in practice, such extensions are unpredictable and staff may not prefer such type of overtime. Thus, there is an implicit inconvenience cost associated with the overtime cost that is not considered in our work. Similarly, we do not consider the inconvenience costs associated with an anesthesiologist being on call, but not being asked to come in to work. While these aspects can be included in our model by suitably appending the overtime and call costs with the appropriate inconvenience costs, quantifying these costs would be challenging. In this regard, recent research in structural estimation (Olivares et al. 2008) could potentially be used to calculate these inconvenience costs and further enhance the outputs of the model. Finally, some anesthesiologists can be used across multiple specialties and this feature was incorporated in our model. However, we do not consider their preferences across specialties as such data was unavailable to us.

Future work could focus on all these aspects to improve the model and its ability to attend to the interests of the surgical teams.

In conclusion, the methodology described in this paper has had a major economic and organizational impact at the operating service department at the UCLA RRMC. This organization expects to maintain the described gains and to increase them continuously several years into the future.

Acknowledgments

References

AMA (2013) http://www.ama-assn.org/ama/pub/news/news/2012-09-17-cpt-code-changes-2013.page.

Augusto V, Xie X, Perdomo V (2010) Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers & Industrial Engineering* 58(2):231–238.

Bates DW, Gawande AA (2003) Improving safety with information technology. *New England journal of medicine* 348(25):2526–2534.

Batun S, Denton BT, Huschka TR, Schaefer AJ (2011) Operating room pooling and parallel surgery processing under uncertainty. *INFORMS J. Comput.* 23(2):220–237.

Beliën J, Demeulemeester E (2008) A branch-and-price approach for integrating nurse and surgery scheduling. *Eur. J. Oper. Res.* 189(3):652–668.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press).

Bertsimas D, Gupta V, Kallus N (2013) Data-driven robust optimization. *arXiv preprint arXiv:1401.0212* .

Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1):35–53.

Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. *Oper. Res.* 54(1):150–168.

Birge JR, Louveaux FV (1988) A multicut algorithm for two-stage stochastic linear programs. *Eur. J. Oper. Res.* 34(3):384–392.

Cardoen B, Demeulemeester E, Beliën J (2009a) Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics* 119(2):354–366.

Cardoen B, Demeulemeester E, Beliën J (2009b) Sequencing surgical cases in a day-care environment: an exact branch-and-price approach. *Computers & Operations Research* 36(9):2660–2669.

Cardoen B, Demeulemeester E, Van der Hoeven J (2010) On the use of planning models in the operating theatre: results of a survey in flanders. *The International journal of health planning and management* 25(4):400–414.

Carøe CC, Schultz R (1999) Dual decomposition in stochastic integer programming. *Operations Research Letters* 24(1):37–45.

Cima RR, Brown MJ, Hebl, et al. (2011) Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *Journal of the American College of Surgeons* 213(1):83–92.

Denton B, Gupta D (2003) A sequential bounding approach for optimal appointment scheduling. *IIE Transactions* 35(11):1003–1016.

Denton B, Viapiano J, Vogl A (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science* 10(1):13–24.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. *Oper. Res.* 58(4-part-1):802–816.

Dexter F, Traub RD (2002) How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesthesia & Analgesia* 94(4):933–942.

Dexter F, Wachtel RE (2014) Strategies for net cost reductions with the expanded role and expertise of anesthesiologists in the perioperative surgical home. *Anesthesia & Analgesia* 118(5):1062–1071.

Dexter F, Wachtel RE, Epstein RH (2016) Decreasing the hours that anesthesiologists and nurse anesthetists work late by making decisions to reduce the hours of over-utilized operating room time. *Anesthesia & Analgesia* 122(3):831–842.

Gawande A (2010) *Complications: A surgeon's notes on an imperfect science* (Profile Books).

Ghaly RF (2014) Do neurosurgeons need neuroanesthesiologists? should every neurosurgical case be done by a neuroanesthesiologist? *Surgical Neurology International* 5.

Green LV, Savin S (2008) Reducing delays for medical appointments: A queueing approach. *Oper. Res.* 56(6):1526–1538.

Gul S, Denton BT, Fowler JW, Huschka T (2011) Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations management* 20(3):406–417.

Gupta D (2007) Surgical suites' operations management. *Production and Operations Management* 16(6):689–700.

Gurobi Optimization I (2015) Gurobi optimizer reference manual, http://www.gurobi.com.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) *Multivariate Data Analysis*, volume 6 (Pearson Prentice Hall Upper Saddle River, NJ).

Haynes AB, Weiser TG, Berry WR, Lipsitz, et al. (2009) A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine* 360(5):491–499.

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.

Hosseini N, Hallbeck MS, Jankowski CJ, Huddleston JM, Kanwar A, Pasupathy KS (2014) Effect of obesity and clinical factors on pre-incision time: Study of operating room workflow. *AMIA Annual Symposium Proceedings*, volume 2014, 691 (American Medical Informatics Association).

Kayis E, Wang H, Patel M, Gonzalez T, Jain, et al. (2012) Improving prediction of surgery duration using operational and temporal factors. *AMIA Annual Symposium Proceedings*, volume 2012, 456 (American Medical Informatics Association).

Kelley JE Jr (1960) The cutting-plane method for solving convex programs. *Journal of the Society for Industrial & Applied Mathematics* 8(4):703–712.

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

Koenker R (2005) *Quantile Regression.* Number 38 (Cambridge university press).

Koenker R (2013) Quantreg: Quantile regression. *R package version* 5.

Kougias P, Tiwari V, Orcutt S, Chen A, Pisimisis G, Barshes NR, Bechara CF, Berger DH (2012) Derivation and out-of-sample validation of a modeling system to predict length of surgery. *The American Journal of Surgery* 204(5):563–568.

Laskin DM, Abubaker AO, Strauss RA (2013) Accuracy of predicting the duration of a surgical operation. *Journal of Oral and Maxillofacial Surgery* 71(2):446–447.

Macario A (2010) What does one minute of operating room time cost? *Journal of Clinical Anesthesia* 22(4):233–236.

Mak HY, Rong Y, Zhang J (2014a) Appointment scheduling with limited distributional information. *Management Sci.* .

Mak HY, Rong Y, Zhang J (2014b) Sequencing appointments for service systems using inventory approximations. *Manufacturing Service Oper. Management* 16(2):251–262.

Mancilla C, Storer R (2012) A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions* 44(8):655–670.

Marcon E, Dexter F (2006) Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science* 9(1):87–98.

Märkert A, Gollmer R (2008) User's guide to ddsip–AC package for the dual decomposition of two-stage stochastic programs with mixed-integer recourse.

Marques I, Captivo ME, Pato M Vaz (2014) Scheduling elective surgeries in a portuguese hospital using a genetic heuristic. *Operations Research for Health Care* 3(2):59–72.

Martin RK (1999) *Large Scale Linear and Integer Optimization: A Unified Approach* (Springer Science & Business Media).

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. *Anesthesia & Analgesia* 103(6):1499–1516.

McNicol R (1997) Paediatric anaesthesia–who should do it? the view from the specialist hospital. *Anaesthesia* 52(6):513–515.

Meskens N, Duvivier D, Hanset A (2013) Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems* 55(2):650–659.

Min D, Yih Y (2010) Scheduling elective surgery under uncertainty and downstream capacity constraints. *Eur. J. Oper. Res.* 206(3):642–652.

Neuhauser D, Provost L, Bergman B (2011) The meaning of variation to healthcare managers, clinical and health-services researchers, and individual patients. *BMJ quality & safety* 20(Suppl 1):i36–i40.

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

Orkin FK, Mcginnis SL, Forte GJ, Peterson, et al. (2013) United States anesthesiologists over 50: retirement decision making and workforce implications. *Survey of Anesthesiology* 57(2):101–102.

Pardo M (2014) The development of education in anesthesia in the United States. Fagerberg J, Mowery D, Nelson R, eds., *The Wondrous Story of Anesthesia* (Springer).

Roxanne Moster hhsshransbiusnas (2014) .

Ruszczyński AP (2006) *Nonlinear Optimization*, volume 13 (Princeton university press).

Saadouli H, Jerbi B, Dammak A, Masmoudi L, Bouaziz A (2015) A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers & Industrial Engineering* 80:72–79.

Sahinidis NV (2014) *BARON 14.3.1: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual.

Thiele A, Terry T, Epelman M (2009) Robust linear optimization with recourse. *Rapport Technique* 4–37.

Tulabandhula T, Rudin C (2014) Robust optimization using machine learning for uncertainty sets. *arXiv preprint arXiv:1407.1097* .

van Rossum F G Drake (2001) Python reference manual, pythonlabs, virginia, USA, http://www.python.org.

Viswanathan J, Grossmann IE (1990) A combined penalty function and outer-approximation method for minlp optimization. *Computers & Chemical Engineering* 14(7):769–782.

Table 1   Summary of resource by specialty

Surgery Specialty	Number of ORs Available	Number of Anesthesiologists Available
Vascular	1	9
Neuro	3	10
Plastics	23	NA
ENT	23	NA
Urology	23	NA
Liver	1	8
Thoracic	2	5
Cardiac	3	14
Trauma	1	NA
Pediatric	2	12
Eye Surgery	23	NA
General	23	NA

Table 2   Surgery characteristics provided in data sets

Surgery characteristics	Description	Variable Name
Realized surgery duration	In hours	ACTUALHRS
Surgeon's estimate of surgery duration	In hours	BOOKEDHRS
Patient Class	Inpatient, Outpatient or Same Day Admit	PATCLASS
Booked Current Procedural Terminology (CPT) code	Medical code maintained by American Medical Association defines the services to be performed during surgery. A surgery may have multiple CPT codes. The surgeon provides a list of services that maybe performed as a part of the surgery. The realized CPT codes may and often do vary from the booked CPT code. Surgeries in our data set covered 2700 unique CPT codes.	CPT
ASA Score	A system for assessing fitness of patients before surgery, higher number signifies a less fit patient. Takes integer values between 1 and 6.	ASA
Patient age	In years	AGE
Surgery Service	Cardiac Surgery, Neuro Surgery, etc., full list as in Table EC.1	SERVICE
Surgeon's name	Names of 493 surgeons unique surgeons (providers) who have performed surgeries in the period over which data was available	PROV
Number of CPT codes	Number of CPT codes associated with procedure	NUMCPT

Table 3 $\bar{\mathcal{W}}(\rho, \tau')$ across ρ and τ'

$\rho \diagdown \tau$	0.1	0.2	0.3	0.4
0.8	1.52	1.43	1.48	1.57
0.85	1.43	1.35	1.45	1.59
0.9	1.41	1.27	1.44	1.59
0.95	1.38	1.24	1.4	1.55
0.98	1.38	1.29	1.42	1.57

Table 4   Data sets for performance analysis

	Number of surgeries conducted per day		Number of anesthesiologists working per day		Number of ORs functioning per day	
	Weekdays	Weekends	Weekdays	Weekends	Weekdays	Weekends
Minimum	30	1	28	1	4	1
Maximum	62	15	38	14	23	11
Average	42	6	32	5	22	6
95 percentile	53	11	36	8	23	9

Table 5   Problems used for performance analysis

Instance	No. of surgeries, $|I|$	No. of rooms, $|R|$	No. of anesthesiologists, $|A|$
A	10	3	5
B	15	5	8
C	25	7	10
D	40	10	25
E	65	23	40

Table 6   Performance Evaluation of Heuristic Procedures for small scale problems

Instance	c_q	c_{oa}	c_{or}	% change in cost of Model Based Heuristic from SAA solution	% change in cost of Practitioner's heuristic solution from SAA solution
A	1000	150	450	0	2.97
	1000	150	900	0	4.61
	1000	150	225	0	1.15
	1000	300	450	0	2.34
	1000	75	450	0	0.75
	2000	150	450	0	2.97
	500	150	450	0	2.97
B	1000	150	450	-1.15	3.64
	1000	150	900	-1.74	5.67
	1000	150	225	0	1.35
	1000	300	450	-0.74	2.76
	1000	75	450	0	0.73
	2000	150	450	-1.15	3.64
	500	150	450	-1.15	3.64

Table 7   Performance Evaluation of Heuristic Procedures for medium and large-scale problems

Instance	c_q	c_{oa}	c_{or}	% change in cost of Model Based Heuristic from Practitioner's heuristic solution
C	1000	150	450	-4.55
	1000	150	900	-6.45
	1000	150	225	-2.56
	1000	300	450	-3.46
	1000	75	450	-2.26
	2000	150	450	-5.55
	500	150	450	-4.57
D	1000	150	450	-6.64
	1000	150	900	-4.45
	1000	150	225	-3.37
	1000	300	450	-5.52
	1000	75	450	-2.65
	2000	150	450	-7.55
	500	150	450	-4.75
E	1000	150	450	-7.03
	1000	150	900	-5.74
	1000	150	225	-4.47
	1000	300	450	-6.69
	1000	75	450	-3.45
	2000	150	450	-7.56
	500	150	450	-4.75

**Table 8**   Results from historical validation

| Surgeries per day	% of days	% reduction of cost of Model Based Heuristic from cost of actual plan	Cost savings ($)
<10	28	0	
10-30	4	2.4	
30-40	18	3.3	
40-50	41	7.2	
50-65	9	8.9	

Table 9   Results from live validation

Surgeries per day	% of days	% reduction of cost of Model Based Heuristic from cost of actual plan	Cost savings ($)
<10	30	0	
10-30	0	0	
30-40	11	2.1	
40-50	48	6.4	
50-65	11	9.1	

Table 10   Summary of Results Before and After Implementation of Decision Support System

Attributes	Before	After	% Reduction
Average number of anesthesiologists on call per day	6.0	5.6	6.7
Average overtime per day for anesthesiologists (hours)	18.2	17.5	3.7
Average daily utilization of anesthesiologists (%)	75	77.6	-3.5
Average number of operating rooms used per day	20.4	18.6	8.6
Average overtime per day for operating rooms (hours)	18.5	18	2.7
Average utilization of operating rooms per day (%)	78	81	-3.8
Average daily operating room costs ($)	57,350	52,417	8.6
Average daily overtime costs ($)	2,2375	2,1754	2.8
Average daily call costs ($)	7,145	6,527	8.5
Average total daily costs ($)	86,870	80,729	7.1

Figure 1   Histogram of average daily utilization of anesthesiologists
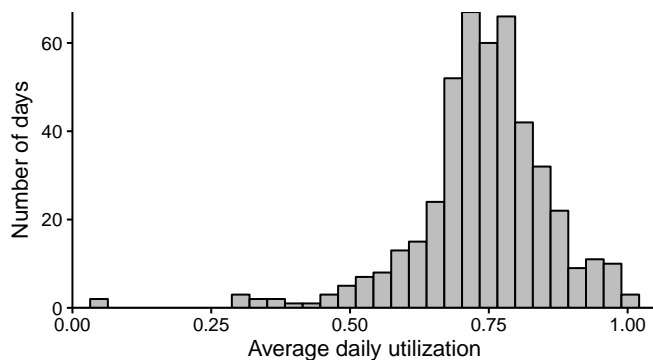
Figure 2   Histogram of average daily utilization of operating rooms
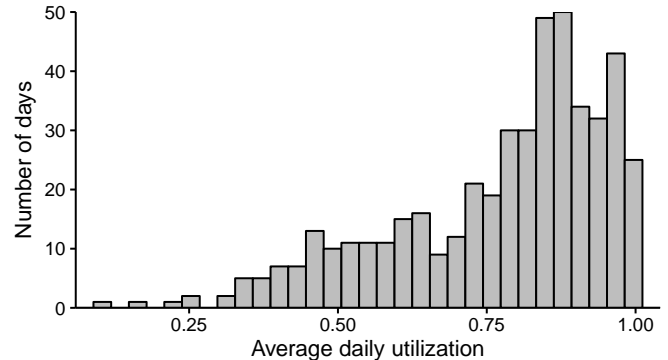
Figure 3   Effect of reducing variability of surgical durations on total resource usage and overtime costs
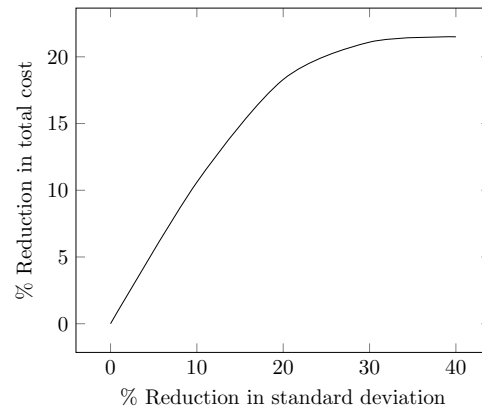
Figure 4   Effect of extending shift timings of ORs on total resource usage and overtime costs
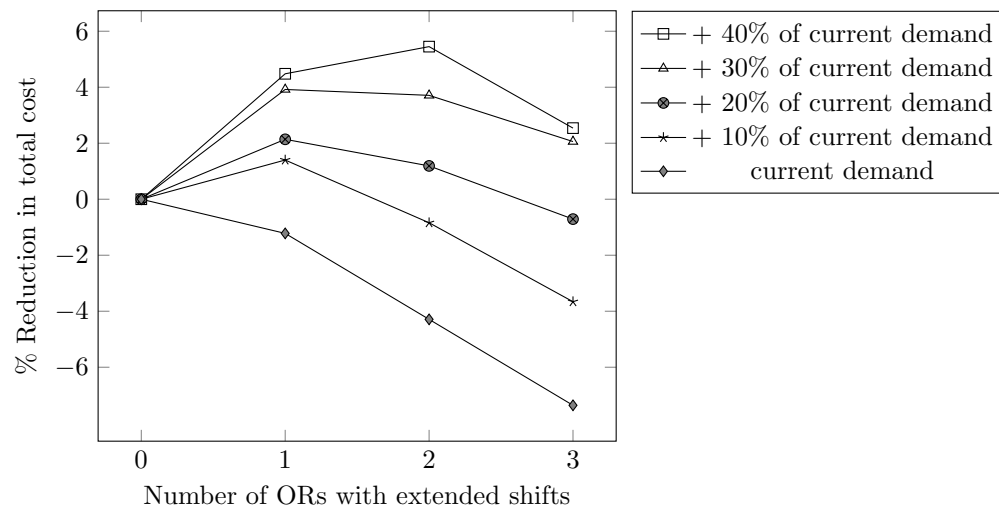
Figure 5   Effect of increasing number of ORs by specialty on total resource usage and overtime costs
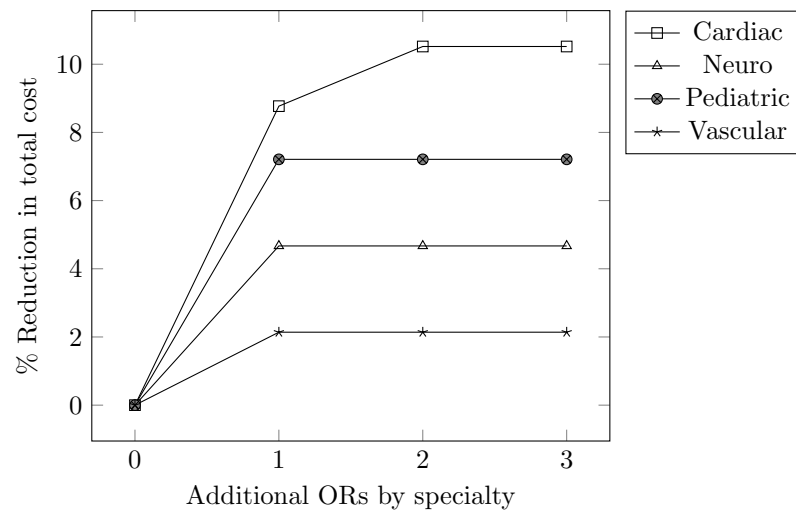
Electronic Companion
Integrated Anesthesiologist and Room Scheduling for Surgeries: Methodology and Application

Sandeep Rath
Kenan-Flagler Business School
University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599
sandeep@unc.edu

Kumar Rajaram
UCLA Anderson School of Management, Los Angeles CA 90095
kumar.rajaram@anderson.ucla.edu

Aman Mahajan
Department of Anesthesiology and Perioperative Medicine
David Geffen School of Medicine at UCLA, Los Angeles CA 90095
amahajan@mednet.ucla.edu

EC.1. The Assignment and Scheduling Decision Making Environment

Assignment and scheduling decisions at operating services department in the UCLA RRMC is complicated by the large number of Operating Rooms (ORs) and anesthesiologists, variety in surgical procedures, variability in anesthesiologist workload, and unpredictability in surgical durations. This section provides more details on these aspects. Table EC.1 shows the average number of procedures performed per day and the range in the number of surgeries across the various specialties. Figure EC.1 shows the uncertainty in the total anesthesiologist workload per day. The two most apparent sources of this variability are in the number of surgeries and the mix of specialties. However, Figure EC.2 shows that even if we control for these factors, there is still considerable variability in total workload per day. This variability arises because of differences across procedures within a specialty and because different patients react differently even to the same procedure (Schaefer et al. 2004). As a direct consequence, the required number of resources (i.e., rooms with anesthesiologists) at any instant as shown in Figure EC.3 and surgical durations are unpredictable. Specifically, the coefficient of variation of surgery duration across these 2700 types procedures varies from 0.75% to 125%.

EC.2. Proof of propositions

Proof of Proposition 1

By definition, the *IARSP* is said to have relatively complete recourse if there exists a feasible second stage solution for any first stage feasible solution (Thiele et al. 2009). Thus, we need to show that there exist feasible $S_i, Over_a, Over_r \forall i \in I, a \in A, r \in R$ satisfying (28) through (32) $\forall \mathbf{d} \in \mathcal{D}(\tau) \ \forall (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1$, where K_1 is the feasibility set of the first stage problem.

Let $\bar{S}_i = \max_{a \in A} t_a^{start} + \sum_{j \in I - \{i\}} \left(\bar{d}_j + \hat{d}_j \right)$ define an upper bound on S_i, the start time of surgery i. This follows as $\bar{S}_i$ would be the start time of surgery i if all surgeries start after the start time of the last shift ($\max_{a \in A} t_a^{start}$), i is the last surgery to be performed in the day after all the other surgeries have

been performed and every other surgery j lasts for its maximum allowed duration $\left(\bar{d}_j + \hat{d}_j\right)$. Thus, S_i is bounded from above. We define $M_{seq} = \max_{i \in I} \bar{S}_i + \max_{i \in I} \left(\bar{d}_j + \hat{d}_j\right)$.

Let $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1$ be some first stage feasible solution. We next show that $S_i = \max_{h \in I} \{s_h\} + \sum_{h \in I, h \neq i} u_{hi} d_h \; \forall i \in I$ is a feasible solution to (28). First observe that by our definition of M_{seq}, this constraint is deactivated when $u_{ij} = 0$. To check the feasibility when $u_{ij} = 1$ note from (12) that $u_{ji} = 0$. Substituting the values of S_i and S_i in (28), we get:

$$\max_{h \in I} \{s_h\} + \sum_{\substack{h \in I \\ h \neq j}} u_{hj} d_h \geq \max_{h \in I} \{s_h\} + \sum_{\substack{h \in I \\ h \neq i}} u_{hi} d_h + d_i - M_{seq} \left(1 - u_{ij}\right) \qquad \forall i, j \in I \qquad \text{(EC.1)}$$

Separating the term $u_{ij} d_i$ from the summation on the LHS and the term $u_{ji} d_j$ from the summation on the RHS leads to:

$$\max_{h \in I} \{s_h\} + \sum_{\substack{h \in I \\ h \neq i,j}} u_{hj} d_h + u_{ij} d_i \geq \max_{h \in I} \{s_h\} + \sum_{\substack{h \in I \\ h \neq i,j}} u_{hi} d_h + d_i + u_{ji} d_j - M_{seq} \left(1 - u_{ij}\right) \quad \forall i, j \in I \quad \text{(EC.2)}$$

Next, setting $u_{ij} = 1, u_{ji} = 0$ and simplifying the above expression we get,

$$\sum_{\substack{h \in I \\ h \neq i,j}} d_h \left(u_{hj} - u_{hi}\right) \geq 0 \qquad \text{(EC.3)}$$

From (13) $u_{hj} \geq u_{hi} + u_{ij} - 1$. As $u_{ij} = 1$, this implies, $u_{hj} - u_{hi} \geq 0$. Therefore, $S_i = \max_{h \in I} \{s_h\} + \sum_{h \in I, h \neq i} u_{hi} d_h \; \forall i \in I$ is a feasible solution to (28) $\forall (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1$. Note that (29) is satisfied since $S_i \geq s_i$. Since there exists a feasible $S_i \; \forall i \in I$ in the recourse problem, then there will always exist $Over_a \geq 0 \; \forall a \in A$ and $Over_r \geq 0 \; \forall r \in R$ for any given $\mathbf{x}, \mathbf{z}$ satisfying (30) through (32). Thus the *IARSP* has relatively complete recourse.

This result can be extended to the case where the integrality condition on $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ is relaxed. This extension will be used in the proof of Proposition 4. In preparation, let $IARSP^{LP}$ be the form of the *IARSP* with the integrality conditions on $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ relaxed and let the feasibility set of $IARSP^{LP}$ be denoted by K_1^{LP}. To prove that $ISRSP^{LP}$ has relatively complete recourse we show by mathematical induction that there exist feasible $S_i \geq 0, Over_a \geq 0, Over_r \geq 0 \; \forall i \in I, a \in A, r \in R \; \forall (\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1^{LP}$ and $\forall \mathbf{d} \in \mathcal{D}(\tau)$. The proof thus far implies that it is now sufficient to prove the existence of a feasible $S_i \forall i \in I$ that satisfies (28). We consider a base case and an induction step, to show this result.

Base case: Let there be two surgeries $I = \{1, 2\}$ and also let there be some $(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1$ and $\mathbf{d} \in \mathcal{D}(\tau)$. Further, let $M_{seq} = \sum_{i \in \{1,2\}} (\bar{d}_i + \hat{d}_i)$. Then the following is a feasible solution to (28) through (32),

$$S_1 = s_1 + s_2 + d_2 + M_{seq}(1 - u_{12}) \qquad \text{(EC.4)}$$

$$S_2 = s_1 + s_2 + M_{seq} \qquad \text{(EC.5)}$$

Proof: Since $u_{12} \leq 1$, (29) is satisfied. For these two surgeries, (28) can be written as,

$$S_2 \geq S_1 + d_1 - M_{seq}(1 - u_{12}) \tag{EC.6}$$

$$S_1 \geq S_2 + d_2 - M_{seq}(1 - u_{21}) \tag{EC.7}$$

Substituting the value of S_1 and S_2 from (EC.4) and (EC.5) in (EC.6),

$$s_1 + s_2 + M_{seq} \geq s_1 + s_2 + d_1 + d_2 + M_{seq}(1 - u_{12}) - M_{seq}(1 - u_{12}) \tag{EC.8}$$

The above simplifies to $M_{seq} \geq d_i + d_2$, which is true by our choice of M_{seq}. Next, substituting the values of S_1 and S_2 in (EC.7),

$$s_1 + s_2 + d_2 + M_{seq}(1 - u_{12}) \geq s_1 + s_2 + M_{seq} + d_2 - M_{seq}(1 - u_{21}) \tag{EC.9}$$

This simplifies to $M_{seq} \geq M_{seq}(u_{12} + u_{21})$ which is true since $u_{12} + u_{21} \leq 1$ from (12). Therefore, the values of S_1 and S_2 satisfy (28) and $IARSP^{LP}$ has relatively complete recourse for $I = \{1, 2\}$.

Induction step: Let $IARSP^{LP}$ have relatively complete recourse for $I = \{1, 2, \ldots, k-1\}$ and for this I, $M_{seq} = M'$ and $S_i' \, \forall i \in I$ be a feasible solution. Thus,

$$S_j' \geq S_i' + d_i - M'(1 - u_{ij}) \qquad \forall i, j \in \{1, 2, \ldots, k-1\} \tag{EC.10}$$

We now prove that $IARSP^{LP}$ has relatively complete recourse for $I = \{1, 2, \ldots, k-1, k\}$.

Proof: Let $I = \{1, 2, \ldots, k\}$, $M_{seq} = 2M'$ and let there be some $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1^{LP}$ and $\mathbf{d} \in \mathcal{D}(\tau)$.

We show that there exists feasible $S_i \, \forall i \in I$. To show this, we first provide a feasible solution for $S_i \, \forall i \in I - \{k\}$. Then we show that for any such feasible $S_i \, \forall i \in I - \{k\}$, there exists a feasible S_k. We define the following feasible solution for $i \in I - \{k\}$

$$S_i = S_i' + d_k + \max_{h \in I}\{s_h\} + M'(1 - u_{ik}) \qquad \forall i \in I - \{k\} \tag{EC.11}$$

All such S_i satisfy (29). To verify that they satisfy (28), we substitute the values of S_i and M_{seq} in (28).

$$S_j' + d_k + \max_{h \in I}\{s_h\} + M'(1 - u_{jk}) \geq S_i' + d_k + \max_{h \in I}\{s_h\} + M'(1 - u_{ik}) + d_i - 2M'(1 - u_{ij}) \tag{EC.12}$$
$$\forall i, j \in \{1, 2, \ldots, k-1\}$$

Simplifying, we get,

$$S_j' \geq S_i' + d_i - M'(1 - u_{ij}) - M'(1 + u_{ik} - u_{jk} - u_{ij}) \tag{EC.13}$$

From (EC.10), $S_j' \geq S_i' + d_i - M'(1 - u_{ij})$. From (13), $1 + u_{ik} - u_{ij} - u_{jk} \geq 0$. Thus the above is always true. Next we show that for all such S_i given by (EC.11) there exists feasible S_k such that,

$$S_k \geq S_i + d_i - M_{seq}(1 - u_{ik}) \qquad \forall i \in I - \{k\} \tag{EC.14}$$

$$S_i \geq S_k + d_k - M_{seq}(1 - u_{ki}) \qquad \forall i \in I - \{k\} \tag{EC.15}$$

$$S_k \geq s_k \tag{EC.16}$$

Substituting the values of S_i from (EC.11) and $M_{seq} = 2M'$ we rewrite the above inequalities to get,

$$S_k \geq S'_i + d_i + d_k + \max_{h \in I}\{s_h\} - M'(1 - u_{ik}) \qquad \forall i \in I - \{k\} \qquad \text{(EC.17)}$$

$$S_k \leq S'_i + \max_{h \in I}\{s_h\} + M'(3 - 2u_{ki} - u_{ik}) \qquad \forall i \in I - \{k\} \qquad \text{(EC.18)}$$

For there to be a feasible S_k, the RHS of (EC.17) must be less than or equal to the RHS of (EC.18) i.e.,

$$S'_i + \max_{h \in I}\{s_h\} + M'(3 - 2u_{ki} - u_{ik}) \geq S'_i + d_k + \max_{h \in I}\{s_h\} + d_i - M'(1 - u_{ik}) \quad \forall i \in I - \{k\} \quad \text{(EC.19)}$$

Simplifying the above we get,

$$M'[4 - 2(u_{ki} + u_{ik})] \geq d_k + d_i \qquad \forall i \in I - \{k\} \qquad \text{(EC.20)}$$

As M' is large enough and $u_{ik} + u_{ki} \leq 1$, the above is true. We also need to check that the RHS of (EC.18) is greater the RHS of (EC.16), i.e.,

$$s_k \leq S'_i + \max_{h \in I}\{s_h\} + M'(3 - 2u_{ki} - u_{ik}) \qquad \forall i \in I - \{k\} \qquad \text{(EC.21)}$$

The above equation is always true as $u_{ik} + u_{ki} \leq 1$. Therefore there exists feasible $S_i \ \forall i \in I$ for $I = \{1, 2, \ldots, k\}$ and by principle of mathematical induction, $IARSP^{LP}$ has relatively complete recourse.

Proof of Proposition 2

For a given first stage solution $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ and an arbitrary $\mathbf{d} \in \mathcal{D}(\tau)$, the dual of the recourse function $\mathcal{R}^{\mathbf{D}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d})$ is given by:

$$\mathcal{R}^{\mathbf{D}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}) = \max \left\{ \sum_{i \in I} d_i \left(\sum_{j \in I - \{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} \right) + \sum_{i \in I} s_i \phi_i - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij}(1 - u_{ij}) \right.$$

$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}(1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}(1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$

subject to,

$$\sum_{i \in I} \mu_{ia} \leq c_{oa} \qquad \forall a \in A \qquad \text{(EC.22)}$$

$$\sum_{i \in I} \theta_{ir} \leq c_{or} \qquad \forall r \in R \qquad \text{(EC.23)}$$

$$\sum_{\substack{j \in I \\ j \neq i}} \lambda_{ij} - \sum_{\substack{j \in I \\ j \neq i}} \lambda_{ji} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} - \phi_i \geq 0 \qquad \forall i \in I \qquad \text{(EC.24)}$$

$$\lambda_{ij}, \mu_{ia}, \theta_{ir}, \phi_i \geq 0 \qquad \forall i, j \in I, a \in A, r \in R \qquad \text{(EC.25)}$$

Here, $\lambda_{ij}, \phi_i, \mu_{ia}, \theta_{ir}, \forall i, j \in I, a \in A, r \in R$ are dual variables corresponding to constraints (28)-(32) respectively. $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ can now be written as

$$\boldsymbol{\mathcal{Q}}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s}) = \max \left\{ \sum_{i \in I} \left(\bar{d}_i + f_i \hat{d}_i \right) \left(\sum_{j \in I-\{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} \right) + \sum_{i \in I} s_i \phi_i - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij} \left(1 - u_{ij} \right) \right.$$

$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} \left(1 - x_{ia} + y_a \right) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} \left(1 - z_{ir} \right) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$

subject to,

$$(EC.22) - (EC.25)$$

$$\sum_{i \in I} |f_i| \leq \tau \tag{EC.26}$$

$$-1 \leq f_i \leq 1 \qquad\qquad \forall i \in I \tag{EC.27}$$

Simplifying the objective function,

$$\boldsymbol{\mathcal{Q}}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s}) = \max \left\{ \sum_{i \in I} \left(\bar{d}_i \pi_i + f_i \pi_i \hat{d}_i \right) + \sum_{i \in I} s_i \phi_i - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij} \left(1 - u_{ij} \right) \right.$$

$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} \left(1 - x_{ia} + y_a \right) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} \left(1 - z_{ir} \right) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$

subject to, $\qquad (EC.22) - (EC.27)$

$$\pi_i = \sum_{j \in I-\{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} \qquad\qquad \forall i \in I \tag{EC.28}$$

$$\pi_i \geq 0 \qquad\qquad \forall i \in I \tag{EC.29}$$

$\boldsymbol{\mathcal{Q}}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s})$ is a bilinear optimization problem because of the presence of bilinear terms $f_i \pi_i$ in the objective function. Further, at optimality $f_i^* \geq 0 \; \forall i \in I$. This is true as $\pi_i \geq 0$ and the feasibility set of f_i is independent of π_i. If τ is a positive integer, constraints (EC.26) and (EC.27) constrain the feasibility set of $f_i \; \forall i \in I$ to be the set $\{0,1\}$. This implies that:

$$f_i \pi_i = \begin{cases} \pi_i, & \text{if } f_i = 1 \\ 0, & \text{if } f_i = 0 \end{cases} \qquad\qquad \forall i \in I$$

We introduce an additional variable $\xi_i = f_i \pi_i$ and rewrite $\boldsymbol{\mathcal{Q}}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s})$ as follows:

$$\boldsymbol{\mathcal{Q}}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{s}) = \max \left\{ \sum_{i \in I} \left(\bar{d}_i \pi_i + \xi_i \hat{d}_i \right) - M_{seq} \sum_{i,j \in I i \neq j} \lambda_{ij} \left(1 - u_{ij} \right) + \sum_{i \in I} s_i \phi_i \right.$$

$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} \left(1 - x_{ia} + y_a \right) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} \left(1 - z_{ir} \right) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$

subject to,

$$(EC.23) - (EC.29)$$

$$\xi_i \le M_f f_i \qquad\qquad \forall i \in I \qquad\qquad (EC.30)$$

$$\xi_i \le \pi_i \qquad\qquad \forall i \in I \qquad\qquad (EC.31)$$

$$\xi_i \ge 0 \qquad\qquad \forall i \in I \qquad\qquad (EC.32)$$

Where, M_f is a sufficiently large positive number. Thus, $\boldsymbol{\mathcal{Q}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ can be written as a mixed-integer program.

Proof of Proposition 3

This proof is adapted from Thiele et al. (2009). For conciseness of notation we represent the set defined by (EC.22)-(EC.25) as $\Lambda(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$. Also let $\boldsymbol{\lambda} = (\lambda_{ij}) \forall i,j \in I$, $\boldsymbol{\mu} = \mu_{ia} \forall i \in I, a \in A$, $\boldsymbol{\theta} = \theta_{ir} \forall i \in I, r \in R$, $\boldsymbol{\phi} = \phi_i \forall i \in I$. From the proof of Proposition 2 and strong duality, we get:

$$\boldsymbol{\mathcal{R}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}) = \max_{\substack{(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\phi}) \\ \in \\ \Lambda(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})}} \left\{ \sum_{i \in I} d_i \left(\sum_{j \in I - \{i\}} \lambda_{ij} + \sum_{a \in A} \mu_{ia} + \sum_{r \in R} \theta_{ir} \right) + \sum_{i \in I} s_i \phi_i - M_{seq} \sum_{i,j \in I i \ne j} \lambda_{ij} (1 - u_{ij}) \right.$$
$$\left. - M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} (1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia} t_a^{end} \right\}$$
$$(EC.33)$$

Let $\boldsymbol{d}^l \in \arg\max_{\boldsymbol{d} \in \mathcal{D}(\tau)} \boldsymbol{\mathcal{R}}(\mathbf{x}^l, \mathbf{y}^l, \mathbf{z}^l, \mathbf{u}^l, \mathbf{s}^l, \mathbf{d})$. From equation (24), $\boldsymbol{\mathcal{Q}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) = \max_{\mathbf{d} \in \mathcal{D}(\tau)} \boldsymbol{\mathcal{R}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d})$. Then for an arbitrary $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$,

$$\boldsymbol{\mathcal{Q}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \ge \boldsymbol{\mathcal{R}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}^l) \qquad\qquad (EC.34)$$

Furthermore, let $(\boldsymbol{\lambda}^l, \boldsymbol{\mu}^l, \boldsymbol{\theta}^l, \boldsymbol{\phi}^l)$ be an optimal solution of (EC.33) with $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}) = (\mathbf{x}^l, \mathbf{y}^l, \mathbf{z}^l, \mathbf{u}^l, \mathbf{s}^l, \mathbf{d}^l)$. From (EC.33) we get:

$$\boldsymbol{\mathcal{R}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}^l) \ge \sum_{i \in I} d_i^l \left(\sum_{j \in I - \{i\}} \lambda_{ij}^l + \sum_{a \in A} \mu_{ia}^l + \sum_{r \in R} \theta_{ir}^l \right) + \sum_{i \in I} s_i \phi_i^l - M_{seq} \sum_{i,j \in I i \ne j} \lambda_{ij}^l (1 - u_{ij})$$
$$- M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l (1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l t_a^{end}$$
$$(EC.35)$$

From, equation (EC.34) and (EC.35) we get,

$$\boldsymbol{\mathcal{Q}}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \ge \sum_{i \in I} d_i^l \left(\sum_{j \in I - \{i\}} \lambda_{ij}^l + \sum_{a \in A} \mu_{ia}^l + \sum_{r \in R} \theta_{ir}^l \right) + \sum_{i \in I} s_i \phi_i^l - M_{seq} \sum_{i,j \in I i \ne j} \lambda_{ij}^l (1 - u_{ij})$$
$$- M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l (1 - x_{ia} + y_a) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l (1 - z_{ir}) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l t_a^{end} \qquad (EC.36)$$

From (EC.28) and (EC.36) we get:

$$
\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \geq \sum_{i \in I} \left(\bar{d}_i \pi_i^l + \xi_i^l \hat{d}_i \right) + \sum_{i \in I} s_i \phi_i^l - M_{seq} \sum_{\substack{i,j \in I i \neq j}} \lambda_{ij}^l \left(1 - u_{ij} \right)
$$

$$
- M_{anesth} \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l \left(1 - x_{ia} + y_a \right) - M_{room} \sum_{\substack{i \in I \\ r \in R}} \theta_{ir}^l \left(1 - z_{ir} \right) - \sum_{\substack{i \in I \\ r \in R}} \theta_{ir} T^{end} - \sum_{\substack{i \in I \\ a \in A}} \mu_{ia}^l t_a^{end} \tag{EC.37}
$$

Therefore, the right hand side of (EC.37) is a lower bound on $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$.

Proof of Proposition 4

The value function of a linear program $z(b) = \min \left\{ c^T x | Ax \geq b \right\}$ is a piecewise linear convex function over the domain for which the linear program is feasible (Martin 1999)[Corollary 2.49, pp. 75].

From the proof of Proposition 1, for any $\mathbf{d} \in \mathcal{D}(\tau)$, if K_1^{LP} is the feasibility set of the first stage problem with the integrality condition on $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ relaxed and $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) \in K_1^{LP}$, then $\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d})$ is a feasible linear program over its domain as $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ appear in the right hand side of its constraints. Thus, from above, $\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d})$ is a piecewise linear convex function in $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ for any given $\mathbf{d} \in \mathcal{D}(\tau)$. From equation (24), $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}) = \max_{\mathbf{d} \in \mathcal{D}(\tau)} \mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d})$. Therefore, $\mathcal{Q}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$ is also piecewise-linear and convex in $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s})$.

EC.3. Sample Average Approximation Solution for the Integrated Anesthesiologist and Room Scheduling Problem

In this section we provide the formulation and a short description of sample average approximation procedure based solution to solve the integrated anesthesiologist and room scheduling problem.

The variable and parameter description is as described in §3 in the paper. The difference between the sample average approximation formulation and the robust optimization formulation is that the surgery duration $d_i \ \forall i \in I$, instead of belonging to an uncertainty set are now modeled as random variables and are denoted by $d_i(\omega)$ under scenario $\omega \in \Omega$. For $|I|$ surgeries the random vector $\mathbf{d}(\omega) = \left\{ d_1(\omega), d_2(\omega), \ldots, d_{|I|}(\omega) \right\}$ is the vector of surgery durations under scenario $\omega \in \Omega$. The support of $\mathbf{d}$ is $\mathbb{R}_+^{|I|}$. Following the evidence found in clinical literature (Strum et al. 2012) and standard assumptions in surgery scheduling literature (Batun et al. 2011, Denton et al. 2010) we assumed that these d_i are independent and have log-normal distribution. The standard formulation of the resulting two-stage stochastic program with recourse is:

$$
\min \left\{ \sum_{r \in R} c_r v_r + \sum_{a \in A} c_q y_a + \mathbb{E} \left[\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}(\omega)) \right] \right\} \tag{EC.38}
$$

subject to,

$$
(2) - (23)
$$

and,

$$
\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}(\omega)) = \min \left\{ \sum_{a \in A} c_{oa} Over_a(\omega) + \sum_{r \in R} c_{or} Over_r(\omega) \right\} \tag{EC.39}
$$

subject to,

$$S_j(\omega) \geq S_i(\omega) + d_i(\omega) - M_{seq}(1 - u_{ij}) \qquad \forall i, j \in I \quad \text{(EC.40)}$$

$$S_i(\omega) \geq s_i \qquad \forall i \in I \quad \text{(EC.41)}$$

$$Over_a(\omega) \geq S_i(\omega) + d_i(\omega) - t_a^{end} - M_{anesth}(1 - x_{ia} + y_a) \qquad \forall i \in I, a \in A \quad \text{(EC.42)}$$

$$Over_r(\omega) \geq S_i(\omega) + d_i(\omega) - T^{end} - M_{room}(1 - z_{ir}) \qquad \forall i \in I, r \in R \quad \text{(EC.43)}$$

$$S_i(\omega), Over_a(\omega), Over_r(\omega) \geq 0 \qquad \forall i \in I, a \in A, r \in R \quad \text{(EC.44)}$$

Assuming a log-normal distribution, we estimate the conditional mean and standard deviation for each cluster of CPT codes. From this estimated distribution, we draw $N_s = 1000$ samples and formulate the sample average approximation of the two-stage stochastic program as:

$$\min \left\{ \sum_{r \in R} c_r + \sum_{a \in A} c_q y_a + \frac{1}{N_s} \sum_{n=1}^{N_s} \mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}^n) \right\} \qquad \text{(EC.45)}$$

subject to,

$$(2) - (23)$$

and,

$$\mathcal{R}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{s}, \mathbf{d}^n) = \min \left\{ \sum_{a \in A} c_{oa} Over_a^n + \sum_{r \in R} c_{or} Over_r^n \right\} \qquad \text{(EC.46)}$$

subject to,

$$S_j^n \geq S_i^n + d_i^n - M_{seq}(1 - u_{ij}) \qquad \forall i, j \in I \qquad \text{(EC.47)}$$

$$S_i^n \geq s_i \qquad \forall i \in I \qquad \text{(EC.48)}$$

$$Over_a^n \geq S_i^n + d_i^n - t_a^{end} - M_{anesth}(1 - x_{ia} + y_a) \qquad \forall i \in I, a \in A \qquad \text{(EC.49)}$$

$$Over_r^n \geq S_i^n + d_i^n - T^{end} - M_{room}(1 - z_{ir}) \qquad \forall i \in I, r \in R \qquad \text{(EC.50)}$$

$$S_i^n, Over_a^n, Over_r^n \geq 0 \qquad \forall i \in I, a \in A, r \in R \qquad \text{(EC.51)}$$

We solve the above program by the integer L-shaped decomposition framework (Birge & Louveaux 1997). The results of the solution procedure are provided in Table 6 of the main paper.

EC.4. Quantile Regression Procedure for Estimating Uncertainty Set $\mathcal{D}(\tau)$

We develop a quantile regression model for predicting $g_L(\tilde{\mathbf{b}}; \rho)$ and $g_U(\tilde{\mathbf{b}}; \rho)$, which are used in (45) to estimate the uncertainty set $\mathcal{D}(\tau)$. As the objective of this model was to predict quantiles, we prefer a parsimonious model with predictive power rather than an over-specified model. For evaluation of the predictive power of this model we use an out of sample Mean Square Prediction Error (MSPE) with respect to the conditional median. To avoid over-specification of the model, we use an Akaike Information Criterion (AIC). The dataset Δ^E had 25700 observations. Δ^E was divided into two disjoint

datasets $\Delta^{E-Train}$ (19300 observations) and Δ^{E-Test} (6400 observations). The quantile regression was built on $\Delta^{E-Train}$ and the out-of-sample MSPE was performed on Δ^{E-Test}. The regression model was built using the following steps:

Step 1: Dimension Reduction. Due to the large number of CPT codes and surgeons compared to the number of observations, we performed clustering to reduce the number of factor variables corresponding to CPT codes and surgeons. For surgeries with multiple CPT codes, we concatenated the CPT codes to create a composite code. For example, if a procedure consisted of CPT codes *A, B* and *C*, we created the composite code *A-B-C*. After this procedure the total number of unique surgery types grew from 2706 to 4061 as there are now more unique procedures with each combination being a different code. Subsequently we performed a *k-means* clustering on the median values of observed surgery duration for each unique procedure code. This clustering procedure is similar to that performed in He et al. (2012). We plot the number of clusters against the % variance explained and choose the elbow point for selecting the total number of clusters. We perform similar clustering procedure for the surgeons. From Figure EC.4 and EC.5 we choose 6 clusters for both CPT codes and surgeons. This choice of clusters explained 95% of the variance in median surgical durations. We use these clustered variables in the subsequent quantile regression and name them *CPTCLUST* for the CPT clusters and *PROVCLUST* for the surgeons.

Step 2: Quantile Regression From the dataset Δ^E, there are 8 possible explanatory variables that could be included for modeling surgical durations. However, using the CPT cluster variable (*CPTCLUST*) and surgery service variable (*SERVICE*) at the same time led to an ill-defined model matrix. This was due to the collinearity between these two variables as measured by the Variance Inflation Factor (VIF) as described in Hair et al. (2006). The VIF for *SERVICE* was 40.19. We removed this variable and found that none of the other variables demonstrated a $VIF > 10$, thus meeting the criteria for not exhibiting significant collinearity (Table EC.2).

Subsequently we performed quantile regression incorporating the above 7 variables. We then sequentially remove variables in increasing order of significance and compare the out-of-sample MSPE and the AIC for the resulting models. The results are shown in Table EC.3. Incorporating surgeon information (i.e., *PROVCLUST*), CPT information (i.e., *CPTCLUST*), age and ASA score improves the prediction and leads to a lower AIC value. Therefore, these variables were incorporated in the model. The number of CPT codes (NUMCPT) was excluded from the final model specification as adding it did not lead to any improvement in MSPE or AIC. The final specification of the quantile regression model was thus,

$$ACTUALHRS = BOOKEDHRS + ASA + AGE + PATCLASS + CPTCLUST + PROVCLUST$$
$$\text{(EC.52)}$$

Next, in Table EC.4 we provide the coefficients of the above quantile regression model at several quantile values (10th, 30th, 60th and 90th quantile). This table demonstrates an advantage of using quantile

regression instead of OLS regression. In particular, it allows for the effect of variable to be different at different quantiles. The coefficient of BOOKEDHRS, representing the surgeons' estimate of surgery duration, increases with increasing quantiles of surgery duration. This implies that more weight is provided to surgeons estimate for longer surgeries. This could imply that surgeons provide a more accurate estimate for longer surgeries than for shorter surgeries. One reason for this could be that surgeons often round up or down to the nearest quarter of an hour while providing estimates. This rounding off would lead to more significant differences for shorter surgeries than for longer surgeries. Another variable that demonstrates changing coefficients with quantiles is ASA. The coefficient of the ASA code increases with increasing quantiles. This could imply that for longer surgeries each increase in ASA score contributes more to the surgery duration than for shorter surgeries. This is also intuitive as the fitness of the patient before surgery would be more significant for longer and more complex procedures.

EC.5. The Decision Support System at the UCLA RRMC

The decision support system for integrated anesthesiologist and operating room scheduling at the UCLA RRMC was built using the Python programming language and a schematic of this system is shown in Figure EC.6. This system consists of an uncertainty set estimation module and an optimization module. The surgery characteristics were provided using the CareConnect database at the hospital and inputted to the uncertainty set estimation module. The output from this module was sent along with anesthesiologist availability, and the surgery resource specialty information provided by the Qgenda database to the optimization module. This solved the IARSP using the model based heuristic and generated an optimized schedule for the following day. This specified the assignment decisions of anesthesiologists and operating room to surgeries along with their scheduled start times. The output of this module was provided to the planner who made adjustments as needed to accommodate special requests by surgeons to change start times or for rooms with additional specialized equipment or for specific anesthesiologists.

There were several challenges in implementing this system. First, the CareConnect database and the Qgenda database had to be accessed daily, and their output had to be reformatted to be compatible with the uncertainty set estimation and the optimization modules. This necessitated the development of a specialized automated interface which required regular maintenance. Second, the planners were initially skeptical about the ability of our system to consider all specialties and constraints. Further, they were unsure if the prediction of surgical durations was better than estimates made by the surgeons, and if the model had adequately captured uncertainty in surgical durations. They felt that if these aspects were not effectively incorporated, this could lead to schedule disruptions, unsatisfied patients and extensive overtime costs. To ensure that the planners were confident with this system, we ran the model in parallel with their approach as described in §5.2 of the paper, so that they could understand the solution of the model and compare this with their own rules. They were reassured that the model solution corresponded

to their solution when the number of daily surgeries was small. However, they also appreciated how the model solution outperformed their rules when the number of daily surgeries was high and the resulting scheduling complexity was larger. In this scenario, the model was more effective in utilizing the flexibility of the multiple parallel resources and considering uncertainty in surgical durations. A third challenge in implementing the system was that surgical characteristics had to be updated and new procedures added to the system. Since these required specialized clinical input, a formalized procedure had to be instituted, where the feedback of the surgeons and anesthesiologists had to be solicited and manually updated in the CareConnect database. This was time consuming and had to be done in monthly basis. However, this was essential to ensure the continued efficacy of our system.

EC.6. Evaluation of the Prediction and Scheduling Benefits of the Model Based Heuristic

In this section, we analyze what proportion of the gains were due to a better prediction method for surgical durations and how much was due to the scheduling policy. To conduct this analysis, we used the estimates of surgical durations made by the surgeons and ran our model for the live and historical validation. Note that the resulting gains would now be entirely from the scheduling policy. We then subtracted these gains from the original benefits which considered prediction and scheduling to calculate the gains of the prediction method. This analysis showed that on average 41% of the benefits were due to better prediction and 59% was due a better scheduling policy. These results demonstrate that due to the complexity of the problem, the most benefits can be got by combining prediction and optimization, an aspect for which robust optimization is particularly well suited.

References

Batun S, Denton BT, Huschka TR, Schaefer AJ. 2011. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS J. Comput.* 23(2) 220-237.

Birge, JR, Louveaux F. 1997. *Introduction to Stochastic Programming.* Springer, New York.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR. 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Oper. Res.* 58(4-part-1) 802816.

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) $99 - 114$.

Schaefer, AJ, Bailey MD, Shechter SM, and Roberts MS. "Modeling medical treatment using Markov decision processes." *Operations research and health care*, pp. 593-612. Springer US, 2005.

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

Table EC.1   Number of surgeries across specialties

Surgery Specialty	Average Number of Surgeries per day	Range of surgeries per day
Vascular	1.75	0-10
Neuro	5.97	0-18
Plastics	6.01	0-15
ENT	1.53	0-6
Urology	3.23	0-15
Liver	5.15	0-10
Thoracic	1.66	0-12
Cardiac	7.07	1-12
Trauma	1.02	0-7
Pediatric	1.34	0-6
Eye Surgery	1.44	0-5
General	26.34	0-62

Table EC.2   Variance Inflation Factors

Variable	VIF
BOOKEDHRS	3.22
ASA	1.23
AGE	1.15
PATCLASS	1.76
CPTCLUST	4.96
PROVCLUST	3.19
NUMCPT	1.04

Table EC.3　　Results of independent variable selection for quantile regression

	Dependent variable:				
	ACTUALHRS				
	(1)	(2)	(3)	(4)	(5)
BOOKEDHRS	0.206***	0.206***	0.204***	0.206***	0.210***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
ASA	0.044***	0.044***	0.036***		
	(0.009)	(0.009)	(0.009)		
AGE	−0.001***	−0.001***			
	(0.0002)	(0.0003)			
NUMCPT	−0.001*				
	(0.0003)				
CPTCLUST2	−4.966***	−4.964***	−4.954***	−5.001***	−4.955***
	(0.145)	(0.147)	(0.157)	(0.143)	(0.133)
CPTCLUST3	−5.872***	−5.869***	−5.869***	−5.908***	−5.888***
	(0.147)	(0.149)	(0.159)	(0.145)	(0.135)
CPTCLUST4	−6.622***	−6.619***	−6.616***	−6.662***	−6.652***
	(0.148)	(0.150)	(0.160)	(0.146)	(0.136)
CPTCLUST5	−3.726***	−3.725***	−3.720***	−3.768***	−3.724***
	(0.147)	(0.149)	(0.159)	(0.145)	(0.134)
CPTCLUST6	−2.350***	−2.353***	−2.342***	−2.373***	−2.336***
	(0.147)	(0.150)	(0.159)	(0.146)	(0.137)
PROVCLUST2	0.363***	0.367***	0.354***	0.374***	0.333***
	(0.066)	(0.067)	(0.067)	(0.068)	(0.065)
PROVCLUST3	0.967***	0.970***	0.962***	0.953***	0.969***
	(0.094)	(0.096)	(0.094)	(0.092)	(0.095)
PROVCLUST4	0.484***	0.483***	0.478***	0.475***	0.464***
	(0.024)	(0.025)	(0.024)	(0.025)	(0.025)
PROVCLUST5	0.331***	0.330***	0.332***	0.331***	0.329***
	(0.015)	(0.016)	(0.016)	(0.017)	(0.018)
PROVCLUST6	0.417***	0.416***	0.400***	0.399***	0.398***
	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)
PATCLASS-INPATIENT	0.077**	0.074**	0.078**	0.078***	
	(0.032)	(0.032)	(0.033)	(0.029)	
PATCLASS-SAME DAY ADMIT	0.159***	0.159***	0.151***	0.141***	
	(0.031)	(0.031)	(0.032)	(0.029)	
Constant	7.245***	7.241***	7.218***	7.359***	7.421***
	(0.156)	(0.158)	(0.168)	(0.153)	(0.141)
Observations	19,335	19,335	19,335	19,335	19,335
AIC	58529	58528	58549	58563	58589
Out-of-sample MSPE	1.312	1.312	1.313	1.313	1.314

*p<0.1; **p<0.05; ***p<0.01

Table EC.4　　Coefficients of quantile regressions at various quantile levels

|	*Dependent variable:*			
	ACTUALHRS			
	$q = 0.1$	$q = 0.3$	$q = 0.6$	$q = 0.9$
BOOKEDHRS	0.096***	0.151***	0.243***	0.469***
	(0.017)	(0.009)	(0.010)	(0.020)
ASA	−0.023***	0.011	0.066***	0.122***
	(0.009)	(0.009)	(0.010)	(0.023)
AGE	−0.002***	−0.002***	−0.001***	−0.0004
	(0.0003)	(0.0002)	(0.0003)	(0.001)
CPTCLUST2	−4.337***	−5.027***	−4.895***	−4.563***
	(0.684)	(0.099)	(0.121)	(0.214)
CPTCLUST3	−4.843***	−5.919***	−5.803***	−5.143***
	(0.683)	(0.101)	(0.122)	(0.220)
CPTCLUST4	−5.254***	−6.573***	−6.591***	−6.219***
	(0.684)	(0.103)	(0.124)	(0.225)
CPTCLUST5	−3.420***	−3.783***	−3.688***	−3.433***
	(0.693)	(0.107)	(0.125)	(0.216)
CPTCLUST6	−2.821***	−2.495***	−2.328***	−2.420***
	(0.701)	(0.115)	(0.122)	(0.216)
PROVCLUST2	0.318***	0.297***	0.473***	0.924***
	(0.064)	(0.079)	(0.077)	(0.123)
PROVCLUST3	0.391***	0.742***	1.087***	1.812***
	(0.134)	(0.096)	(0.083)	(0.193)
PROVCLUST4	0.228***	0.414***	0.520***	0.764***
	(0.039)	(0.025)	(0.029)	(0.066)
PROVCLUST5	0.207***	0.295***	0.325***	0.380***
	(0.026)	(0.018)	(0.017)	(0.047)
PROVCLUST6	0.181***	0.352***	0.417***	0.538***
	(0.032)	(0.021)	(0.020)	(0.050)
PATCLASS-INPATIENT	0.185***	0.075	0.080	0.176
	(0.071)	(0.078)	(0.066)	(0.142)
PATCLASS-SAME DAY ADMIT	0.589***	0.274***	0.093	−0.141
	(0.075)	(0.078)	(0.065)	(0.141)
Constant	5.543***	7.130***	7.237***	7.005***
	(0.691)	(0.136)	(0.146)	(0.282)
Observations	19,335	19,335	19,335	19,335
Note:			*p<0.1; **p<0.05; ***p<0.01

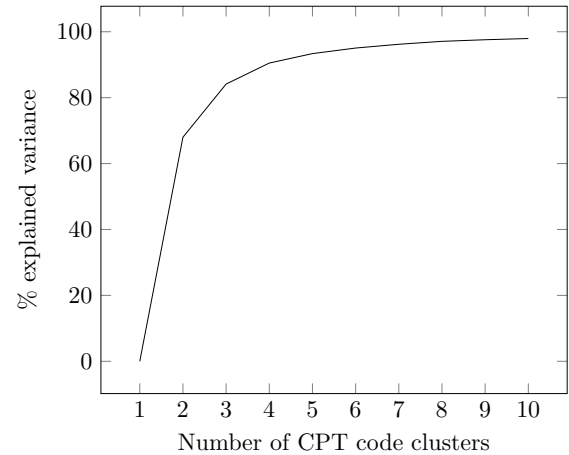**Figure EC.1**   Histogram of average daily anesthesia hours
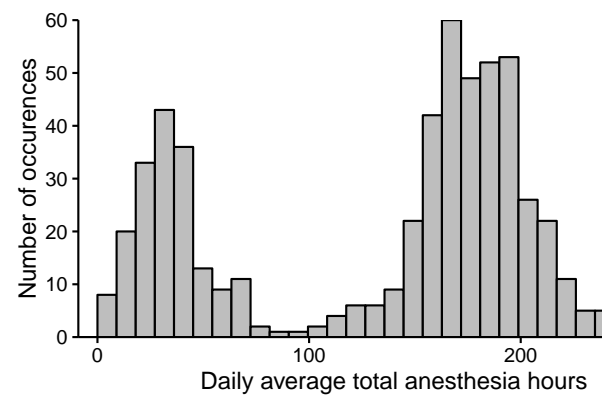
Figure EC.2   Histogram of average daily anesthesia hours controlling for number and mix of surgeries
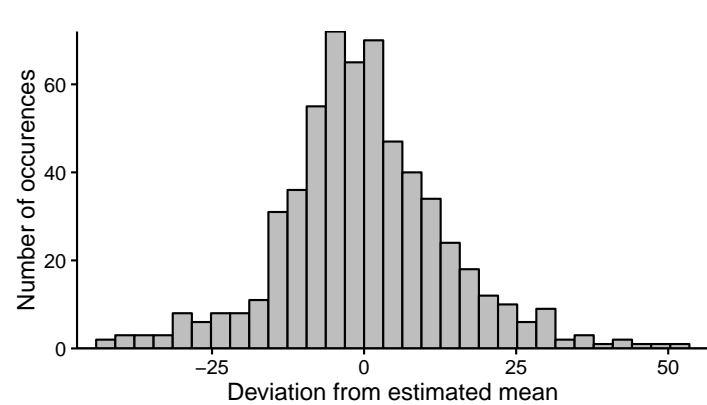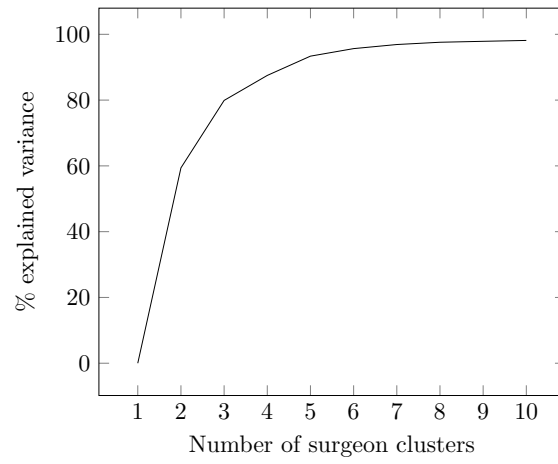

Figure EC.3   Number of rooms with assigned anesthesiologists by hour of day
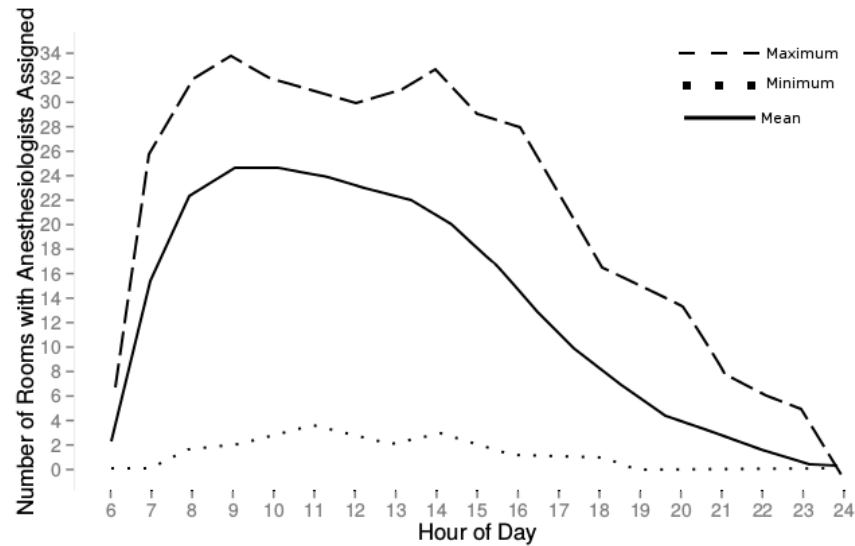
Figure EC.4   Number of CPT code clusters and % explained variance

Figure EC.5   Number of surgeon clusters and % explained variance

Figure EC.6   Software schematic of decision support system