Clemson University
TigerPrints

All Theses | Theses

8-2014

The Impact of Block Scheduling and Release Time on Operating Room Efficiency

Rebecca Weiss
Clemson University, rhweiss@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Part of the Industrial Engineering Commons

Recommended Citation
Weiss, Rebecca, "The Impact of Block Scheduling and Release Time on Operating Room Efficiency" (2014). *All Theses*. 1875.
https://tigerprints.clemson.edu/all_theses/1875

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

THE IMPACT OF BLOCK SCHEDULING AND RELEASE TIME ON OPERATING ROOM EFFICIENCY

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Industrial Engineering

---

by
Rebecca Weiss
August 2014

Accepted by:
Dr. Kevin Taaffe, Committee Chair
Dr. Lawrence Fredendall
Dr. Amin Khademi

ABSTRACT

Planning for sufficient surgical capacity at a hospital requires that many tactical and operational decisions be made before the day of surgery. Typically, blocks of time in operating rooms (ORs) are assigned and specific surgical cases are placed in rooms. The hospital monitors utilization to determine the schedule's effectiveness in balancing the risk of overtime with idle time. In this thesis, we will examine how adjusting schedule risk ratios and penalty values, and providing shared, open posting time affected the hospital's ability to identify an efficient but high quality and low cost block schedule. The proposed schedules were tested by assigning surgical cases to ORs and simulating the schedule's performance using recent data from a local hospital. We also show how scheduling accuracy can impact the performance level of the schedules proposed. Once the schedule has been set, the use of block release time is investigated in order to provide insight on how to better fill these ORs and increase utilization levels. Release policies are simulated based on various surgery arrival distributions, capacity levels, and case durations. We will show how different policies involving assigned and open posting rooms impact utilization levels, number of cases not fit into the schedule, and number of cases posted after the block release time.

DEDICATION

This thesis is dedicated to my partner, Anh Thu, who has supported me throughout this entire process by encouraging me to get involved with the research community and inspiring me to pursue my passions, my parents, Jacinda and Stephen, who push me to do well in school, my brother, Mitchell, who takes interest in my work, my Auntie Chris, who is proud of me and my accomplishments, Pol, who is always supportive, even when he's halfway around the world, my friends, for reassuring and standing by me, and all of my wonderful teachers, who have inspired me to do great things and taught me both material and life lessons along the way.

ACKNOWLEDGMENTS

I would like to thank Dr. Taaffe, Dr. Fredendall, and Dr. Khademi for being on my committee and giving me insight about my thesis and research. I am extremely grateful to Dr. Taaffe for giving me the wonderful opportunity to dive into this field of research and for his patience when guiding me throughout the process of proposing research questions, writing papers, and encouraging my work on this thesis.

I would like to acknowledge and thank Charlie Hajzus and the Medical University of South Carolina for the real-world insights, suggestions, and approaches and for the data used in Chapters I and III.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

CHAPTER I
INTRODUCTION

In the field of healthcare, methods to improve process flow and efficiency have become increasingly important as hospitals look for ways to decrease costs without affecting the quality of care. The Operating Room, or OR, is the most costly resource in the hospital, causing it to be a large focus of the healthcare literature [3]. This thesis will contribute to the current healthcare literature by focusing on how to effectively schedule and manage a set of ORs in order to keep desired utilization levels and reduce the number of unscheduled cases.

Before a surgery occurs in an OR, there are several steps that must be taken to coordinate when, where, and how the surgery will take place. First, a Master Schedule is created by allocating time to surgical groups based on historical data and forecasting methods. Hours can be assigned to particular surgical groups, referred to as blocked time, or assigned to all groups as an open or flexible room, referred to as open posting time. OR schedule creation has been studied in the literature with focuses on minimizing cost, maximizing profit, obtaining high utilization levels, and reducing overtime [3,4,5,6,7,8,9,16,17]. Next, a Master Surgical Schedule is created by assigning surgeries to particular rooms at different times. The process of surgical scheduling has been extensively studied in the literature as a means of increasing efficiency, reducing overtime, and reducing the number of unscheduled cases each day [18,19,20,21,22,23,24]. Any unscheduled time within the assigned blocks may then be released, or made available to other surgical groups, to allow for excess and add-on cases to be added to the schedule. The effects of different block release policies are not common in the healthcare literature, but current research shows that having block release does not significantly OR efficiency and utilization levels [30,32]. Despite these results, in practice, block release policies are commonly used [33].

The research in this thesis will focus on the allocation of hours to ORs and determination of block release policies for those ORs. In order to properly assign hours and rooms to surgical groups, we present mathematical models that use a flexible risk tolerance ratio and open posting encouragement value to account for differences in environment and preferences. By allowing for flexibility in the allocation of OR time, OR managers can allocate time based on expected growth, utilization level, or willingness to allow overtime. We show the impacts of different averseness ratio and encouragement levels on OR utilization and number of unscheduled cases. Once blocks have been assigned, we examine block release policies for proposed surgical groups with different surgery arrival rates, surgical durations, and expected capacities. Finally, we perform a case study to combine both of these studies. We use data from a local hospital to create a Master Schedule, simulate surgery cases arriving, assigning cases to rooms, and then playing out the day of surgery. We note that the research does not focus on the exact sequence and assigned time of surgeries, although we do make note of and use methods of assigning surgeries to one room over another in the literature [18,32]. This work differs from the current literature because of the flexible scheduling tools it provides and the in-depth analysis of block release policies that have not been heavily studied.

The thesis will continue as follows: Chapter II will present mathematical models used to allocate hours to ORs and assign those hours to specific rooms. Chapter III will use simulation to suggest block release times based on surgery arrival rates, case durations, and expected room utilizations. Chapter IV discusses a case study that combines the work of Chapters II and III by using historical data to create multiple OR Master Schedules and then using these schedules to suggest block release times for surgical groups.

CHAPTER II

BALANCING RISK IN OPERATING ROOM SCHEDULING

1. INTRODUCTION

Improving operating room (OR) efficiency and utilization has become a strong priority for hospitals in recent years. Hospitals are looking to reduce cost without reducing quality of care [1]. The OR schedule affects not only costs and OR utilization, but also surgeon, staff, and patient satisfaction. Each hospital has different resource constraints, policy limitations, and variance in case scheduling accuracy, so the complexity of reducing inefficiencies can vary greatly from hospital to hospital. Past surgery data can be used to simulate possible outcomes and costs from different schedule options in order to find a schedule that will best suit the hospital [2,3,4,5,6,7,8,9,10].

In this research, we describe the models used to allocate surgical time to service groups, and then play out scenarios using recent booked and actual lengths of surgery data. We propose two mathematical models that provide easy-to-use mechanisms that allow for flexibility and a range of options to explore when setting reserved block sizes within a suite of ORs. We further show how adjustments to OR and block allocation policies can affect critical hospital performance measures, such as scheduled hours, rooms, and utilization levels.

2. LITERATURE REVIEW

When creating a new OR block schedule, there are three general steps to consider: (1) forecast of demand based on prior usage and expected changes (new surgeons or predicted monthly caseload changes), (2) allocation of time blocks to surgical groups and staff scheduling, and (3) assignment of time blocks to specific ORs.

Mathematical modeling is a common approach for optimally allocating time to surgeons, surgical groups, and wards based on historical data. Objectives for this type of problem range from minimizing cost to maximizing profit to multi-criteria objective functions. McIntosh et al. [2] stress the importance of refining OR allocations two to three months prior to the day of surgery. Cardoen et al. [3] surveyed hospitals in Belgium and found that the majority focused their objectives on high utilization of ORs and avoidance of overtime. Strum et al. [4,5] use a cost analysis model involving an overtime-to-undertime risk ratio and a distribution of caseload hours to pinpoint the number of block hours required to minimize expected cost. Similarly, Hosseini and Taaffe [6] use an overtime-to-undertime risk ratio to minimize cost by finding a balance between the number of surgical hours needed from week to week. Blake and Donald [7] use integer programming to find optimal block allocations for surgical groups and use penalties in the objective function if the desired number of hours are not assigned. Dexter et al. [8] discuss ways of allocating surgical time based on resource utilization and general hospital profitability and rewarding surgical groups that require less ward and ICU time. Kuo et al. [9] recommend allocating block hours using data about service revenue and efficiency levels.

Researchers have also considered allocation of blocks based on the daily surgery queue. Mannino et al. [10] propose a model that minimizes queue costs based on a finite number of resources. Dexter and Macario [11] compare three methods: one that fixes a service's block hours, one that allows surgeons to post cases on any work day, and one that allocates OR hours based on a time limit for the patient to receive care. They argue that in a fixed block hour system, specialties that have the greatest caseload variance are penalized, but if cases are scheduled on a weekly basis per the surgeons' request, this could improve efficiency and allow for fewer cases to be turned away.

Simulation is another technique used to improve block allocation and surgical scheduling as well as validate updates to the current OR settings. Kumar and Shim [12] use simulation techniques to model process changes for a hospital in Singapore. Their simulation exposed bottlenecks in the system's pre-operation area. Ballard and Kuhl [13] found the surgical suite's maximum capacity without using additional resources or overtime. They also explained the downfalls of underestimating the OR's utilization level. Dexter et al. [14] showed that increasing utilization past 90% does not have a substantial impact on overall utilization and revenue. This will most likely cause a decrease in revenue for the hospital.

While total block allocation is often the focus of many studies, it is not necessarily clear how to place this time into ORs in the existing suite. ORs may vary in size and equipment capability, meaning not all rooms should be treated equally. Surgeons and surgical groups also have requirements and preferences that need to be accommodated in the block allocation. Testi et al. [15] develop a formula to maximize surgeon preferences with capacity, volume, and allocated time constraints. They also construct an algorithm to assign patients to each room. Shamayleh et al. [16] use an objective function to simultaneously allocate time blocks, decide which ORs to open, and manage overtime hours. Denton et al. [17] consider the deterministic and stochastic OR allocation problems through minimization of opening and overtime OR costs. Blake and Donald [7] match OR hours with target block hours given the number of ORs needed per surgical specialty. While these studies are similar to our work, they contain some unrealistic assumptions, such as infinite capacity, lack of room restrictions, equal operational lengths, and weekly solving of the block schedule. We propose a model that takes a more realistic approach in addressing these issues by building in flexibility for OR managers, creating room and capacity restrictions, and allowing for less frequent solving of the block schedule.

Once the master surgical schedule is set, surgeries are scheduled into specific ORs. Dexter and Traub [18] found that the earliest start time rule, scheduling a new case into the OR that will be opened the earliest, performs well to minimize overutilized OR hours. Denton et al. [19] found that scheduling surgeries in increasing order of duration variance produced the optimal result more frequently than other heuristics. Van Houdenhoven et al. [20] showed that a cyclic OR surgical scheduling policy increased utilization, decreased the number of cancelled cases, and allowed for a more leveled patient outflow for the Intensive Care Unit. Agnetis et al. [21] focus on creating a weekly Master Surgical Schedule with elective cases based on characteristics such as surgical duration, waiting time, and priority level. Van Houdenhoven et al. [22] found that scheduling surgeries using mathematical algorithms and relaxed organizational policies showed increases in OR utilization levels. Lehtonen et al. [23] found that planning for one OR team to work longer increased OR efficiency and productivity. Stepaniak et al. [24] found differences in surgical scheduling methods between risk and nonrisk averse OR coordinators. Van Essen et al. [24] consider the rescheduling surgeries problem. They developed a Decision Support System that provides three adjusted OR schedules when case times must be adjusted. Their results showed that rescheduling the OR mostly leads to either shifting a surgery's start time or providing a scheduled break between two consecutive surgeries.

In our research, we provide a new framework of mixed integer programs that address the allocation of block time and ORs as well as summarize the performance of the block allocations by simulating the posting of cases to the rooms. The "simulation" is carried out within an MIP that attempts to complete the caseload by efficiently using the existing blocks and ORs (which will be explained in later sections). Our approach uses recent surgery data to simulate outcomes about idle, on time, and overtime ORs to show differences in the proposed block schedules. Finally, we note that our MIP framework does not replace the need for an intelligent method for sequencing or scheduling surgeries. In this research, we focus on the problem of block allocation and block assignment to ORs.

3. METHODS, ASSUMPTION, AND APPROACH

Data across 24 months of surgery postings was obtained from a level one trauma, teaching hospital in the Southeastern U.S. At the time, the hospital had 26 different service specialties, 3 locations, and 39 ORs. For each surgical case, the data extracted was: procedure date, patient in and out of room time, booking start and end times, location, booking room, surgical service, and associated case level.

We used IBM's ILOG CPLEX Optimizer to determine schedules, room placement, and surgery set allocations. OR and service-specific constraints, as well as location constraints, were included in the models. Surgery lengths were considered to be time of patient arrival to time of patient departure. Surgeries that had a case level of 1A, 1B, 2, or 3 were considered to be part of the Urgent or Urgent/Emergent block time. Surgeries that had a case level of 4 or 5 were considered to be part of the service-specialty block time. All surgery types were considered, and all surgeries were allowed to use open posting block time. No surgeries starting before 6 AM or after 6 PM were included. In order that we account for an accurate turnover time between cases, a turnover time analysis was conducted by location by subtracting the previous surgery's procedure end time from the next surgery's procedure start time. Any length that was greater than 90 minutes was excluded from the analysis. The average turnover time was added to all surgical cases except the final surgery of the day in an OR.

We use twelve months of historical surgery data when creating optimal block hours for each surgical specialty, and we reserve the next three most recent months of historical data for analyzing actual surgery posting to the blocks. Epstein & Dexter [26] found that using nine to twelve months of historical data is best for staffing solutions. When using nine months of data, we found inconsistencies between quarters and seasonality issues that were misguiding the schedule. For this reason, we chose to proceed with using twelve months of data. Resulting utilization levels of the schedule were calculated by dividing the total amount of surgery hours used, including turnover time, by the total number of reserved block hours on the schedule. We used a total of 24 months of historical data for the analysis with 12 months used for setting up the initial blocks and 12 months, broken down into four quarters, for testing performance. This is described further in section 4.

In the next subsections, we will discuss the mathematical models used to set and test the block schedule. Figure 1.1 depicts the flow of steps and the relationships of the models to the analyzing and evaluating portions of this research. The first step, setting the block schedule, contains the two models that are the focus of this paper. These models use historical surgery data, a risk ratio factor, and an encouragement for open posting time factor to specify hours and rooms for each surgical specialty. The models have been separated for ease of use and personalization. Depending on the hospital’s goals, one or both models can be used. The benefit of having the models split like this will be further explained in Section 4.6. The second step, measuring performance, uses a more recent set of surgery data reserved to simulate a day of surgery with the proposed block schedule.

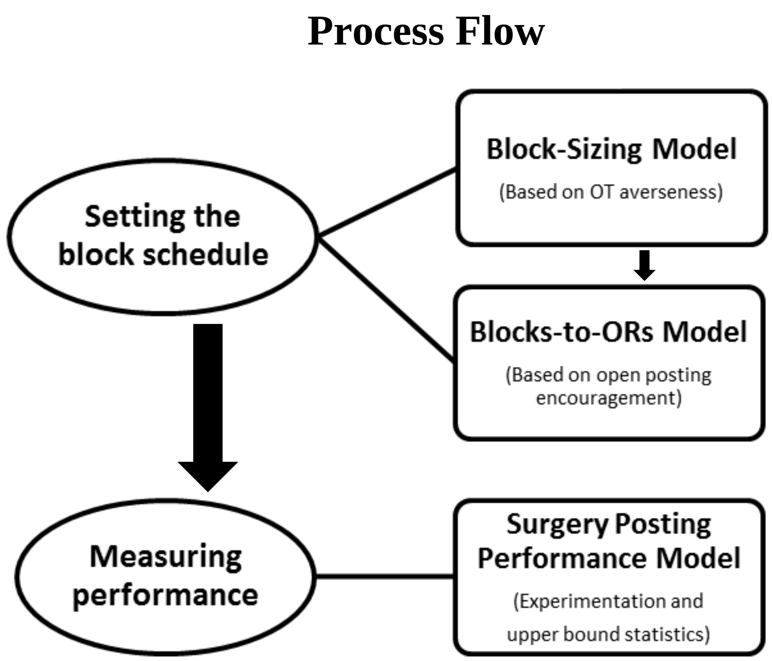

Figure 1.1: The flow of models used to set and test the block schedule

3.1 Determining an OR Block Allocation Plan

3.1.1 Block Sizing Model

The first step to create an OR block schedule is to allocate block hours to each surgical group based on prior demand, as well as projected increases or decreases. As seen in Figure 1.1, this relates to the "Setting the block schedule" portion of the process. We build upon a previous integer programming (IP) model suggested by Hosseini and Taaffe [6] that uses historical data and an overtime-to-undertime cost ratio to find an optimal number of block hours for each service specialty. This ratio can be interpreted as a "level of averseness" to overtime, and it can be adjusted for different groups or locations as needed, which allows for more or less protection of assigned block hours. After hours are assigned or allocated to the services, an open posting block is created for any groups that did not receive a block of time, as well as for groups that may occasionally request more time than originally assigned.

Hours were assigned based on allowable block hour combinations. We allowed for block sizes of 8, 10, and 12 hours, as used by the hospital from which we acquired the data. Two of the three locations were allowed to have a maximum block size of 12 hours, and the third location required a maximum block size of 10 hours. The MIP, Block-sizing Model (BSM), is as follows:

Model Notation:

Subscripts

i index for surgical groups; from 1 to I (groups in total)

j index for weeks of historical data; from 1 to J (weeks in total)

d index for days of the week; from 1 to 5 with 1 being Monday

Sets

A set of allowable block sizes and combinations

Decision Variables

x_{id} amount of block hours assigned to group i on day d

o_{ijd} amount of overtime incurred by group i during week j on day d

u_{ijd} amount of undertime incurred by group i during week j on day d

Parameters

t_{ijd} hours of surgery time schedule for group i during week j on day d

h_i overtime-to-undertime risk ratio (averseness to overtime) for group i

[BSM]

$$\min \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{d=1}^{5} \left(u_{ijd} + h_i * o_{ijd}\right)$$

subject to:

$$o_{ijd} \geq t_{ijd} - x_{id} \qquad \forall i, j, d \qquad (1.1)$$

$$u_{ijd} \geq x_{id} - t_{ijd} \qquad \forall i, j, d \qquad (1.2)$$

$$x_{id} \in A \qquad \forall i, d \qquad (1.3)$$

$$o_{ijd}, t_{ijd}, x_{id} \geq 0 \qquad \forall i, j, d. \qquad (1.4)$$

Note that model BSM is separable by day *d*, however for simplicity we write the formulation to cover an entire surgical week. Constraint (1.1) states that overtime is the non-negative portion of total surgery time scheduled less total block hours assigned. Constraint (1.2) states that undertime is the non-negative portion of block hours assigned less total surgery time schedule. Constraint (1.3) ensures that assignments are in the allowable set.

Once the block hours have been allocated, we use the following formula to set hours for the open block time (denoted by *g*):

$$g_d = \sum_{i=1}^{I} \sum_{j=1}^{J} o_{ijd} / J \qquad \forall d. \qquad (1.5)$$

The heuristic shown in Equation (1.5) takes the average overtime of all groups across all weeks of the historical data set and rounds it up to an allowable block size for the surgical schedule. This method allows for groups who were not allocated a block in the BSM to still have time to post cases. It also allows for groups who may occasionally use time over their blocks to post their extra cases somewhere. This is not part of the optimization model but could technically be shown as part of BSM. In other words, variables g_d are a byproduct of the solution generated from BSM. This equation is important because it calculates the allocation of hours to the open posting room.

It is important to note that the model finds an optimal amount of surgical hours based on the risk ratio provided, which, in turn, affects the number of open posting hours provided. The benefits of an open posting block can be seen from two perspectives: cost and unity. In terms of cost, the open posting block does not restrict its surgical time to a specific group. This means that multiple surgical groups can share an OR, depending on the type of surgeries being done, and that there will be less unused block time since there are more options for cases. From a management perspective, an open posting block can assist with providing a sense of unity between surgeons and surgery groups by joining specialties in the same room. When a block is assigned to a particular surgery group, the group is often considered to "own" that block or room. An open posting block would eliminate this sense of ownership and force surgeons to share the time allocated.

3.1.2 Blocks to ORs Model

Once the total number of block hours was assigned to each service specialty, we created another MIP model to sort these hours into ORs. Since this task is related to making schedule assignments, this model is still part of "Setting the block schedule," as seen in Figure 1.1. The objective of this MIP is to minimize the number of rooms used while still giving each specialty group either all or close to their desired number of block hours from the BSM by including an open posting encouragement factor. For example, if Orthopedic were assigned a total of 36 hours, it might be optimal to allocate the service three rooms with 12-hour blocks. If a particular group is not assigned all of the hours they requested from the BSM, the hours they lost are added to the open posting block. However, each hour not given to the group is penalized in the objective. Similar to the averseness to overtime ratio (from the BSM model in Section 3.1.1), this penalty value (denoted as e_i) can also be adjusted. We now present the Blocks to ORs (BTOR) formulation as follows:

Model Notation:

Subscripts

i index for surgical groups; from 1 to I (groups in total)

r index for ORs; from 1 to R (rooms in total)

d index for days of the week; from 1 to 5 with 1 being Monday

l index for number of different facilities; from 1 to L (facilities in total)

Sets

O_l set of open posting groups in I (one per facility)

Decision Variables

x_{ird} number of hours for service group i in room r on day d

y_{rd} binary decision variable denoting if room r on day d is open or not

c_{id} amount of hours not assigned to group i on day d from original, optimal allocation

Parameters

t_{id} optimal total block hours for group i on day d

m_{rd} maximum number of hours allowed in room r on day d

g_{ld} size of open posting block required in location l on day d

e_i encouragement level; penalty for each c_{id} hour that was reallocated for more open posting time

[BTOR]

$$\min \sum_{r=1}^{R}\sum_{d=1}^{5} y_{rd} + \sum_{i=1}^{I}\sum_{d=1}^{5}(e_i * c_{id})$$

subject to:

$$\sum_{i=1}^{I} x_{ird} \le m_{rd} \qquad \forall r, d \qquad (2.1)$$

$$\sum_{r=1}^{R} x_{ird} = t_{id} - c_{id} \qquad \forall i, d \qquad (2.2)$$

$$x_{ird} = 0 \leftrightarrow \text{service can't use room} \qquad \forall i, r, d \qquad (2.3)$$

$$\sum_{i \in O_l}\sum_{r=1}^{R} x_{ird} = g_{ld} + \sum_{i \in O_l} c_{id} \qquad \forall l, d \qquad (2.4)$$

$$y_{rd} \in \{0,1\}; \qquad \forall r, d \qquad (2.5)$$

$$x_{ird}, c_{id} \geq 0 \qquad \forall i, r, d. \qquad (2.6)$$

Constraint (2.1) ensures that the hours assigned do not exceed the maximum number of hours allowed in that room. Constraint (2.2) solves for the variable *c*, the number of hours requested by group *i* that were not filled. Constraint (2.3) ensures that groups that are not allowed to use certain rooms cannot use those rooms. This ensures that room restrictions (groups that have certain room or equipment needs) are followed. Constraint (2.4) uses the unassigned time from constraint (2.2) and adds it to the open posting time.

Constraint (2.4) works as a way to ensure feasibility for hospitals that do not have excess room capacity. While the BSM finds the optimal number of hours for each surgical group, if the hospital does not have the capacity to assign each group their own room to accommodate those needs, the BTOR model selects the best way to reduce hours in certain surgical groups and to create a shared, open posting block. For example, if two surgical groups were assigned 16 hours from the BSM, we can create four 8-hour blocks for both groups to use. However, if the hospital does not have four rooms available, the model can take six hours from each group and make that into a 12-hour, shared, open posting block. This would result in one fewer room, which would help accommodate the capacity constraint. In addition to the capacity constraint, the penalty value is also a deciding factor on when to combine hours and from which surgical group(s). Each time an hour is moved to the open posting block time, the objective incurs a penalty. Changing the penalty value will affect the number of ORs scheduled each day. This value can be a single fixed value for all service groups, or it can be assigned by particular service group to protect their hours from being sent to the open posting block time.

3.2 Surgery Posting Performance Model (SPPM)

Once the service blocks have been adjusted to fit into ORs (via models BSM and BTOR in prior sections), the next step is to use actual case posting data to determine how well the facility can perform on a given day (i.e., the “Measuring performance” portion of the process flow from Figure 1.1). To accomplish this, we use the next three most recent months of historical data. We select three months since block schedules typically do not change more quickly than this in practice.

While our approach in this section is similar to a trace-driven simulation, the platform for the analysis is actually a mixed integer program that we created. Using the booked length of surgery, cases are assigned to either service-specific blocks or open posting blocks. Groups without a service-specific block may only have cases assigned to the open posting blocks. All cases posted are to be fit into the schedule. The goal is to minimize the number of completely full or overbooked rooms, as well as the number of overtime hours. We allow for overtime to occur in case overtime is less costly than opening up a new OR and to ensure that all cases are assigned to a room. We formulate the Surgery Posting Performance Model (SPPM) as follows:

Model Notation:

Subscripts

i index for surgical groups; from 1 to I (groups in total)

r index for OR number; from 1 to R (rooms in total)

k index for the surgery number on a given day; from 1 to K (maximum for any single service on a day)

w index for the week number of the simulation; from 1 to W (weeks in total)

Superscripts

o indicates an open posting room

f indicates a service-specific room

Sets

N set of service-specific ORs

V set of all open posting ORs

C set of all closed ORs

Decision Variables

d^o_{wikr} binary variable denoting if group i's surgery k is scheduled in open posting room r or not

d^f_{wikr} binary variable denoting if group i's surgery k is scheduled in service-specific room r or not

z_{wr} binary variable denoting if room r on week w is full or not

n_{wr} binary variable denoting if room r on week w is used or not

o_{wir} amount of overtime hours incurred by group i on week w in room r

Parameters

q_{wik} booking length of surgery k for group i in week w

b_{ir} size of block (hours) for group i in room r

C_1 fixed cost of a room potentially running into overtime

C_2 cost per hour of having known overtime

[SPPM]

$$\min \sum_{r=1}^{R}\sum_{w=1}^{W}(n_{wr} + C_1 z_{wr}) + C_2 \sum_{i=1}^{I}\sum_{r=1}^{R}\sum_{w=1}^{W} o_{wir}$$

subject to:

$$\sum_{k=1}^{K} q_{wik} d^f_{wikr} \leq b_{ir} * n_{wr} + o_{wir} \qquad \forall w, i, r \in N \qquad (3.1)$$

$$\sum_{k=1}^{K} q_{wik} d^o_{wikr} \leq b_{ir} * n_{wr} + o_{wir} \qquad \forall w, i, r \in V \qquad (3.2)$$

$$\sum_{r=1}^{R} d^f_{wikr} + d^o_{wikr} = 1 \qquad \forall w, i, k \qquad (3.3)$$

$d_{wikr}^{f} = 0 \qquad \forall w, i, k, r \in V, C \qquad (3.4)$

$d_{wikr}^{o} = 0 \qquad \forall w, i, k, r \in N, C \qquad (3.5)$

$o_{wir} \leq b_{ir} * z_{wr} \qquad \forall w, i, r \in N \qquad (3.6)$

$z_{wr} \in \{0,1\}; \qquad \forall w, r \qquad (3.7)$

$n_{wr} \in \{0,1\}; \qquad \forall w, r \qquad (3.8)$

$d_{wikr}^{f}, d_{wikr}^{o}, o_{wir} \geq 0 \qquad \forall w, i, k, r. \qquad (3.9)$

Constraints (3.1) and (3.2) calculate the amount of predicted overtime scheduled. Constraint (3.3) ensures that every surgery is booked. Constraints (3.4) and (3.5) ensure that surgeries are not allowed in closed rooms and surgeries in service-specific (open posting) rooms do not get placed in open posting (service-specific) rooms. Constraint 3.6 ensures that overtime does not occur in a room not being used. Additional constraints can be added to ensure that cases adhere to room and equipment requirements.

Note that constraints can be used to define when a room is "full" when solving for the z_{wr} variable. For example, in order to provide an overtime prevention buffer, the model can consider a room to be full when it has two hours of free space remaining. This free space can later be used for shorter add-on cases or for scheduled cases that run longer than expected or are behind schedule. The model also allows for the user to decide the *C* coefficients based on the final goal. If the goal is to minimize the number of rooms that go beyond their scheduled time, then set $C_1 > C_2 * o_{wir}$. We note that $C_2 * o_{wir}$ represents the total estimated cost of overtime that would be allowed and does not necessarily mean the value of o_{wir} must be known prior to solving. If the goal is to reduce the number of overtime hours and not necessarily the number of rooms with overtime, then set $C_1 \leq C_2 * o_{wir}$. Our analyses below set $C_1 > C_2 * o_{wir}$, making the objective focus more on the number of ORs running beyond their scheduled time.

4. RESULTS AND DISCUSSION

To observe the effect of the new block allocations, once cases are assigned to rooms, the case booking times are then replaced with the actual case length times to simulate a real day of surgery. From there, results such as amount of idle time and number of hours beyond block can be obtained. Note that since the model does not take into account the exact surgery sequence and does not allow for the switching or re-sequencing of cases, results are expected to represent an upper bound in their ability to reduce overtime and overall room requirements. Thus, we would expect better results once in practice.

In Table 1.1, we provide a set of risk ratios, or overtime-to-undertime ratios (as seen in the BSM) to represent a hospital's level of averseness to overtime. In order to represent a wide range of decisions and hospital conditions in our analysis, we provide five key levels of averseness to overtime. In Table 1.2, five levels are also represented for the encouragement of open posting. These levels affect how blocks are put into rooms and how much open posting time is allowed within the schedule. We provide a top penalty value of 0.3 since any greater value would provide no encouragement and produce the exact same results.

Table 1.1: Levels of averseness to overtime

Code	Risk Ratio	Description
Low averseness to OT	1.25	OT is 25% more costly than UT
Low-Medium averseness to OT	1.5	OT is 50% more costly than UT
Medium averseness to OT	1.75	OT is 75% more costly than UT
Medium-High averseness to OT	2.0	OT is 100% more costly than UT
High averseness to OT	2.25	OT is 125% more costly than UT

Table 1.2: Levels of open posting encouragement

Code	Penalty Value	Description
High encouragement (No penalty)	0	Open posting time is highly encouraged
Medium encouragement	0.1	Open posting time is mildly encouraged
Medium-Low encouragement	0.12	Open posting time is mildly encouraged
Low encouragement	0.2	Open posting time is slightly encouraged
Threshold (No encouragement)	0.3	Block hours remain as assigned

4.1 Base Case

We first present our base case results, using a medium averseness and medium-low encouragement. Hosseini and Taaffe [6] found that medium averseness corresponded to an 80% utilization rate for a variety of data sets, which is a desirable rate for many hospitals [26]. According to Green [27], hospitals typically target a utilization level of 85%, however such high utilizations in the operating rooms are typically very difficult to manage. With higher utilization levels comes longer delays and greater patient dissatisfaction. In the hospital where our research was carried out, they would like to target 75%-80% utilization, with blocks being resized if the service approaches 85% utilization. We set the base case penalty to a medium-low encouragement level for the analysis because it represented a compromise between retaining the original blocks as-is and converting an entire block to open time. With this penalty value, the BTOR model will never remove a full block of hours (at least 8) from a room because there is insufficient reward to compensate for this in the objective function.

After feeding 12 months of historical data to the BSM and BTOR models, we fed the next three months in the data set to the SPPM model to place surgical cases into ORs and to simulate case postings to the blocks. This was performed four times (or four quarters) in succession, so that we could accumulate results across one entire year. Without loss of generality, assuming our data set began in January 2012, we used data from January 2012 – December 2012 to create the blocks, then evaluated the actual case postings using January 2013 – March 2013 data. This resulted in Quarter 1 (Q1) results below. Similarly, we used April 2012 – March 2013 data to create blocks, and evaluated actual case postings using April 2013 – June 2013 data. Results for Q1 to Q4 were found in this way. Figure 1.2 displays the base case results from the simulated performance.

It can be noted that, in general, average utilization is between 70% and 80% and that at least 70% of all surgery hours are completed within the booked blocks. It turns out that the surgical demand at our studied facility had dropped off across the following year, which resulted in slightly lower utilizations than what was observed in Hosseini and Taaffe [6]. However, in the second and third quarters, case postings were frequent enough to reach an 80% utilization. The percent of ORs that finish within their blocked time is slightly lower but still between 55%-75%. From an OR manager's perspective, this could mean that if a surgery is planning to go longer than expected, the surgery following it can be moved to an OR that still has time remaining.

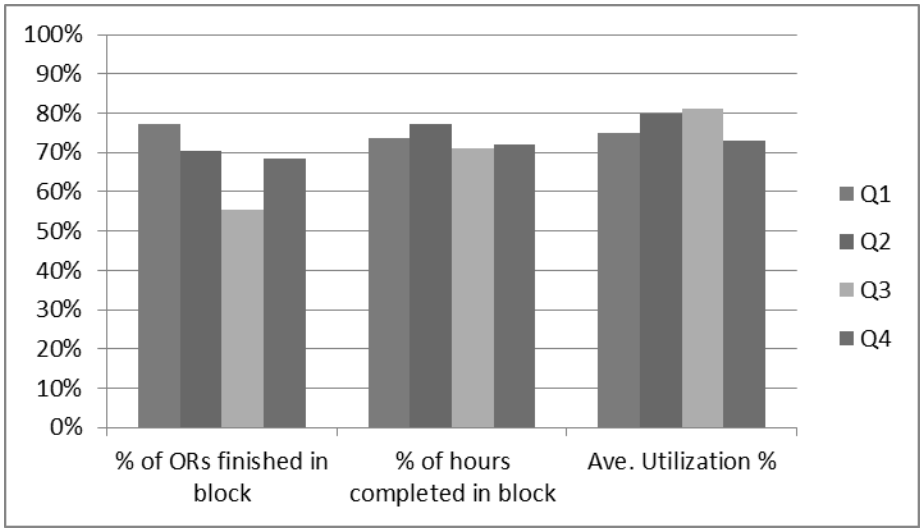

Figure 1.2: Base case results with medium averseness and medium-low encouragement

4.2 Impact of Averseness to Overtime

Figures 1.3 and 1.4 show the impact on the number of reserved block hours and utilization levels in relation to the hospital's averseness to overtime. Note that when the hospital's averseness to overtime is low, overtime is not as costly. As Figure 1.3 shows, this leads to fewer block hours being assigned and the potential for more overtime and higher utilization to occur. On the other hand, using a higher averseness establishes higher overtime costs than idle time costs. With this value, more block hours are assigned, and the possibility of overtime is reduced. However, as a consequence, utilization levels are also reduced. Higher averseness levels can be used as a way to provide buffer between surgeries throughout the day. This buffer can be useful when considering inevitable delays including switching between patients and/or surgical specialties, as well as other delays that may occur while the patient is en route to the operating room. It can also be used if the OR manager is looking to reduce the number of surgery block hours reserved per day to better fit the surgical demand caseload.

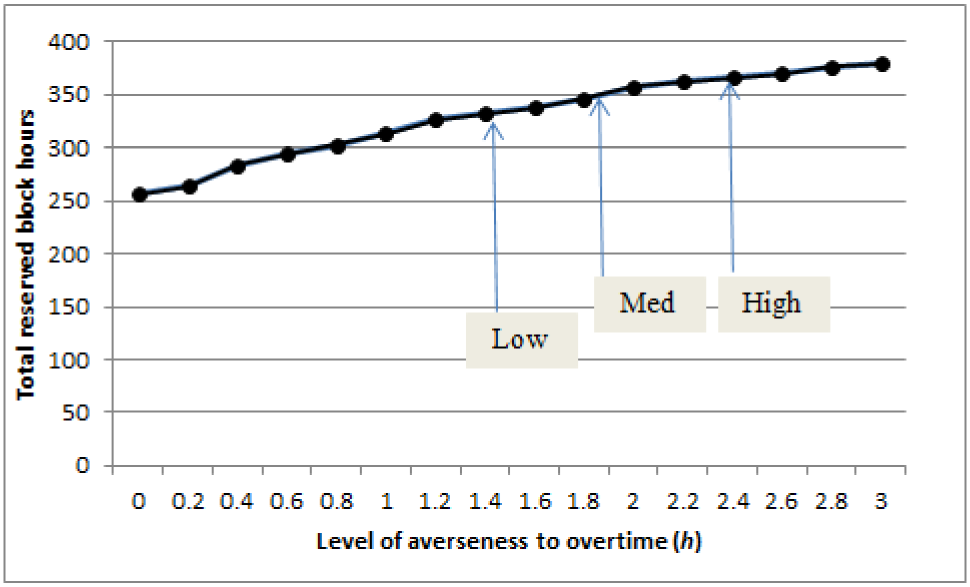

Figure 1.3: Difference in number of reserved block hours for various levels of averseness

Room utilization is an important metric for all hospitals in judging how effective their scheduling policies may be. Figure 1.4 depicts this utilization across the range of 0.0 to 3.0 in the averseness scale (or what is shown as a risk ratio in our models introduced in Section 4.3). This figure is intended to help managers understand where they may fall on their averseness to overtime. Starting at 0.0, where overtime has no cost, the BSM places all hours into open posting blocks.

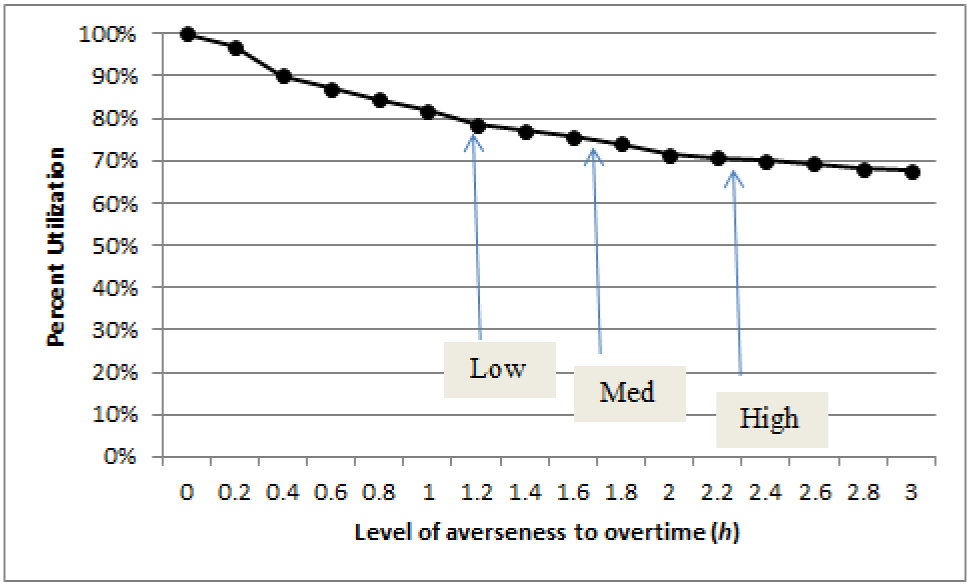

Figure 1.4: Difference in utilization levels for various levels of averseness

It is interesting to note that when shifting from equally balanced overtime and undertime costs (an averseness ratio of 1.0), which would be classified as extremely low averseness, to medium-high averseness (a ratio of 2.0), a 14% increase in the number of reserved block hours and a 10% decrease in utilization levels occurs. *Based on the hospital data used in this analysis*, for every 20% increase (decrease) in the level of averseness to overtime, blocked hours increase (decrease) by 3% and utilization decrease (increase) by 2%. Key results that can be taken from these findings are that having a lower averseness to OT will allow for higher utilization levels and more overtime. Lower averseness levels can be good for surgeries and services with low variability and high accuracy in their scheduling estimates. On the other hand, having a higher averseness to OT will allow lower utilization levels and more undertime to occur. This can be good for surgeries and services that need their hours protected or have surgeries that are highly variable.

We also note that when we reach an averseness level of 3.0, even though each service receives very large reserved blocks, there is still a need for open posting time. In fact, about 6% of the total block time is still allocated for open posting.

4.3 Impact of Open Posting Encouragement

Figure 1.5 illustrates how changing the open posting encouragement level (in the BTOR model) can affect the outcomes presented above. Using different averseness and encouragement levels can provide a number of scheduling options for managers. If an OR manager is looking to reduce the number of ORs scheduled per day, the encouragement level can be adjusted (see Table 1.2). Since any unfilled hours are added to the open posting block, a higher encouragement level (i.e., a lower penalty value e_i) will induce more open posting time because it can group services together and use fewer ORs. However, a lower encouragement level (or higher value of e_i) will help preserve each group's requested block time. In all open posting encouragement tests, utilization will remain constant because, unlike when manipulating the averseness to overtime, the number of total block hours does not change. Instead, the number of rooms may change as block time is shifted between service-specific blocks and open posting blocks, and we can see differences in the number of on-time, early, and late rooms.

We note that, for the studied problem, an encouragement level of 0.3 or greater implies no open posting encouragement. In other words, the penalty for transferring hours becomes too great, and no room adjustments are made. This number may change depending on the number of ORs and possible block hour combinations being used. The number of rooms scheduled based on overtime averseness and open posting encouragement is depicted in Figure 1.5.

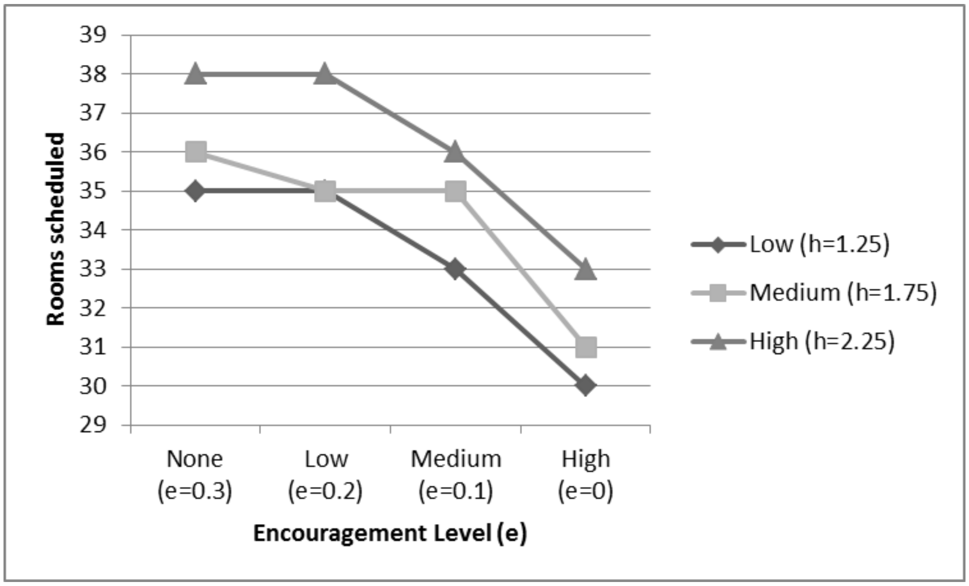

Figure 1.5: Difference in rooms scheduled with combined averseness and encouragement levels

From these results, we observe that shifting from using no open posting encouragement (a threshold level of 0.3) to high open posting encouragement (a level of 0.0) decreased the number of ORs scheduled by 15%. Across encouragement levels, this implies that for every 33% increase in open posting encouragement issued, we see one fewer OR scheduled. Key results to take from this analysis are that high open posting encouragement results in more shared, open posting time on the schedule, while lower open posting encouragement makes little to no changes on the schedule. Low encouragement levels can be used as a starting place if open posting is a new or underused concept.

4.4 Specific Comparisons

We created additional surgical schedules with high and low averseness and encouragement levels to compare to the original base case. These schedules were run through the surgery posting performance model (SPPM) to further show how different emphases affect OR times. The results, shown in Figures 1.6 and 1.7, compare three particular averseness to overtime and encouragement levels and display simulated average performance for the given weekday. In Figure 1.6, low and high levels of averseness to overtime were examined and simulated to compare statistics with the medium averseness, or base case, scenario. A constant medium-low encouragement level was used for the penalty value.

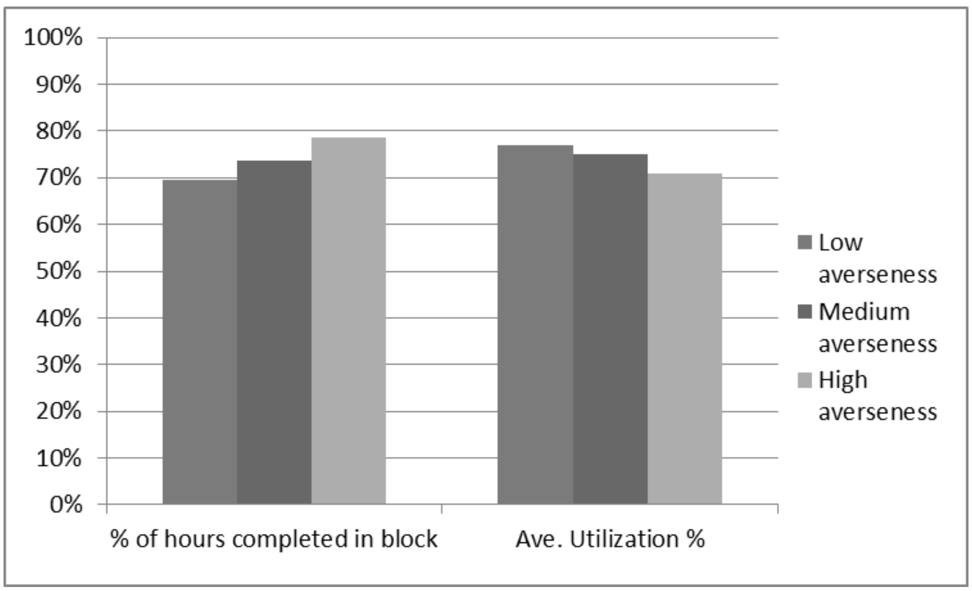

Figure 1.6: Results for three key levels of averseness with medium-low encouragement

Compared to the base case, across the entire week, one to two fewer ORs are assigned per day per quarter using a low averseness. Figure 1.6 shows the average utilization levels, which are calculated by dividing the total number of surgical hours used by the number of hours on the schedule. Utilization levels increase anywhere from 2% to 8%, and we see a slight decrease in number of ORs running on time or early as compared to the base case scenario. Using a high averseness, we see one to two more ORs assigned per day as compared to the base case. We see a decrease in utilization of as much as 10%, although most utilization differences are in the 2-3% range.

In Figure 1.7, no encouragement and high encouragement were compared against the medium-low encouragement level, or base case scenario. A medium averseness-to-overtime of 1.75 was used for all tests.

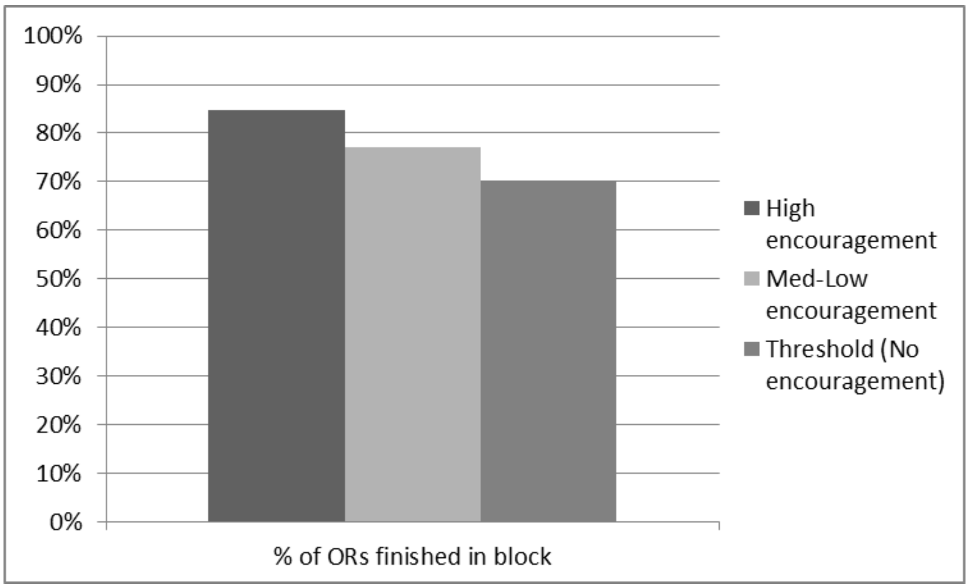

Figure 1.7: Results for three key encouragement levels with medium averseness

Compared to the base case, two to four fewer ORs are assigned using high encouragement. Figure 1.7 shows an increase in the number of ORs that are on time or early due to the increase in the number of open posting rooms available. Using no encouragement, we see one to two more ORs assigned. The number of on-time or early ORs is very close to the base case since the schedules are similar.

4.5 Scheduling Accuracy

While the results in prior sections indicate desirable utilization levels and number of ORs scheduled, there is still room for improvement. This is partly due to the fact that the source data for the experimentation originates from an actual hospital's historical surgeries. Because model SPPM posts cases based on booked length, it does not take into account how long the surgeries actually lasted. Actual results concerning the number of ORs running early, on time, and late rely heavily on the scheduling accuracy of the cases booked. Since we were using historical data, we were able to view the differences between the bookings, but in the real world, only the booking times are known. Figure 1.8 compares the surgical booking time to the actual surgical duration, based on data from the studied hospital. It is easy to see the large differences between the booked hours and actual hours every month.

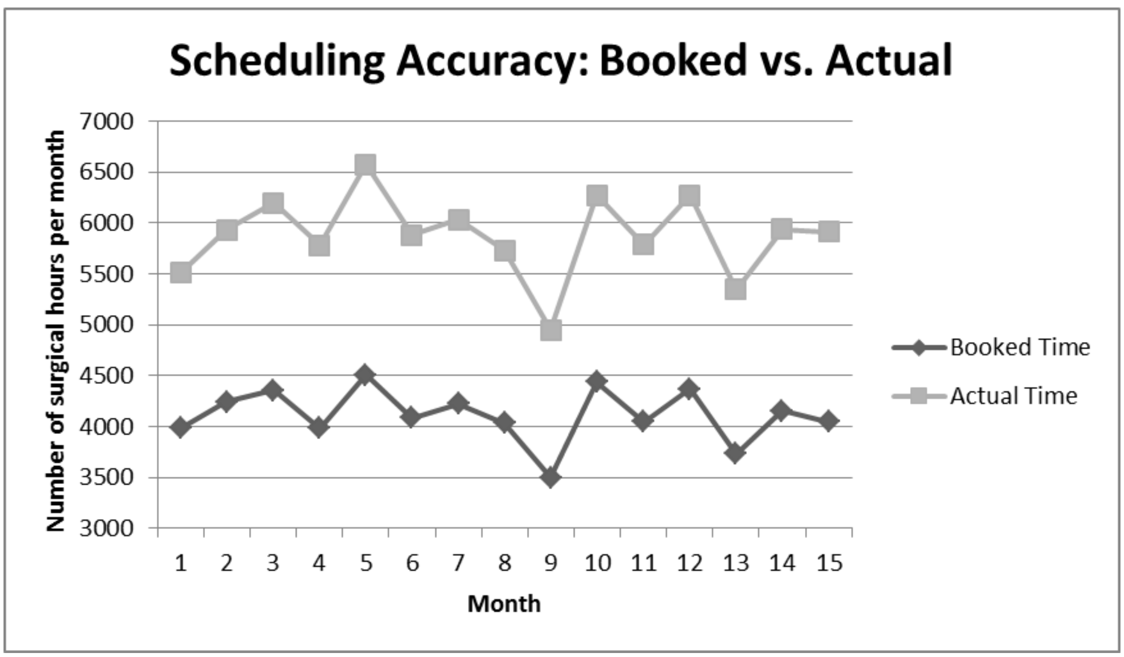

Figure 1.8: Scheduling accuracy: Booked hours vs. actual hours used

This large difference was the motivating factor for testing how improvements in accuracy levels can alter amount of overtime incurred. OR managers can use our analytical modeling approach to encourage greater accuracy in posting cases. Dexter et al. [28] found that case duration estimates should not rely on one single number. They state that certain estimates should rely on the expected mean duration, while others should rely on upper and lower prediction bounds, as well as other uncertainty measures.

We present results showing varying levels of scheduling accuracy and the consequential effects on the amount of early/on time hours and time beyond their reserved block. In Figure 1.9, each level of increased scheduling accuracy indicates that the booked time was closer to the actual length of surgery. At each level of accuracy, the amount over / under booked was decreased by reducing the difference between the two times by that percentage. A 0% increase represents current surgery prediction times, while a 100% increase in accuracy implies that every surgery booking time equaled the actual surgery duration time. The results in Figure 1.9, using the base case assumptions of a medium averseness (h = 1.75) and medium-low encouragement (e = 0.12), show the amount of hours and ORs that complete their surgeries in the allotted block time given based on how accurate their scheduling predictions are. As the predictions of booked time compared to actual time become more accurate, more hours are completed inside the blocked time (less overtime) and more ORs finish on time or early.

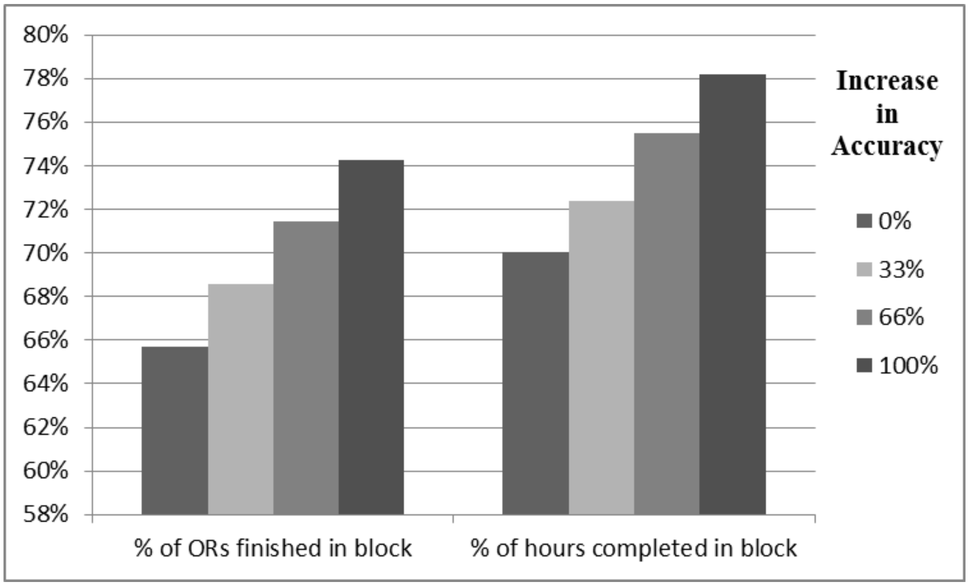

Figure 1.9: Results from scheduling accuracy experiments

Figure 1.9 indicates that even small changes in accuracy levels can have a significant impact on the performance of ORs. Recall that our model plays out the day of surgery based on actual case durations. Since the model assumes that every case must be performed, and even if the case posting times were adjusted to be exactly equal to the actual case durations, then there still could be ORs running into overtime even at 100% accuracy. This would probably not be the case in a real hospital setting because lower priority surgeries would likely be pushed back in order to better fit the schedule. It is also possible that surgeries could be moved into underposted rooms, something not accounted for in the SPPM.

4.6 Managerial Insights

The models proposed in this research provide valuable insight to hospital managers who are making decisions concerning the OR block allocation and case posting processes. These models demonstrate the compromises managers must address about open and reserved blocks as well as risk tolerance levels for overtime to undertime. As noted in Hosseini and Taaffe [6], risk tolerance levels can be especially when exact overtime and undertime costs cannot be quantified. The parameters to use for setting the preferred block allocations will be up to the hospital, although viewing the results from a few block schedule options can allow for new ideas to be pursued.

Many hospitals have policies about flexing out their staff members and closing ORs when not needed. Our method and schedules provide more consistency such that a flexing policy would not be needed. The goal of scheduling surgical hours using an averseness to overtime ratio and a concept of open posting time is to provide a more accurate estimate on the number of hours a surgical group would actually need each week, given that some weeks may be busier than others. Open posting time serves as a way to provide sufficient buffer capacity for handling extra cases in any week.

Because these factors are built into the model, a more constant schedule is maintained throughout the period. If used properly [29], this could improve the attitude of the hospital staff, who might be dissatisfied about being sent home early or not receiving enough hours during the week if ORs are consistently being shut down. Other factors can be built into the model to allow for other management intangibles. For example, uniformity of hours across the week or more manageable block times for nurse schedules can be added to the BSM and BTOR models via additional constraints.

We showed a consistent reduction in number of ORs assigned for this hospital while still maintaining desirable utilization levels. Often times, hospitals are scared to close ORs because of unknown demand, but there is still some level of certainty found within the historical data. By adjusting the level of averseness, OR managers can still decide how aggressive they want to be, depending on the hospital environment, its goals, and the anticipated (or known) amount of variability in caseload. In particular, hospitals with lower variability can employ a lower averseness to greatly reduce the number of ORs assigned per day. Likewise, if a particular service is expected to grow in size, it can be given a higher averseness to overtime ratio so that it receives more surgical hours and has the needed capacity to accommodate the forecasted growth.

Using an averseness to overtime ratio, rather than actual cost numbers, is particularly appealing because of the flexibility it provides. The penalty value (associated with encouragement of open posting) also provides great flexibility in determining the need for an extra OR compared to creating more open posting time. Hospitals with surgical groups that are not used to sharing rooms may want to start off with a very small penalty value so that only a small amount of sharing is encouraged. As groups get more used to the idea, the hospital can encourage more open posting time and potentially require fewer ORs.

The models are flexible in both their parameters and applications. Whether the schedule is being solved weekly or quarterly, and whether the hospital has high or low capacity, this modeling framework can be used to create a less costly and more satisfactory block schedule. The goal was not to instruct hospitals how they should schedule their rooms and block times, but rather to provide options for the management to explore. Factors, conditions, and goals may vary from hospital to hospital, so having a flexible model helps to tailor these methods. Another perk of these tools is that each step is broken down separately. While ideally the steps would be used together, it can be useful to pinpoint certain areas of interest and investigate options involving a particular step. Decision support can be provided for adding or removing block hours and ORs without editing an entire set schedule.

5. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

In this paper, we provided models and methods to create a block allocation schedule that can be tailored to the needs of any large hospital with multiple OR suites. Historical case posting and surgery duration data was used to assess the ability of the block schedule to accommodate the caseload across the available ORs. We showed how changing the overtime-to-undertime risk ratio, or level of averseness, affected the number of rooms used and average utilization levels. In particular, our tests showed that when doubling the risk averseness from equally balanced overtime and undertime costs (i.e., low risk averseness of h=1.0) to overtime being twice as costly (i.e., h=2.0), we observed a 14% increase in the number of reserved block hours and a 10% decrease in utilization levels. Over several settings, we observed that for every 20% increase in averseness to overtime, blocked hours will increase by 3% and utilization will decrease by 2%. While these exact results come from the test hospital used, in general, we expect to see an increase in the number of reserved block hours and a decrease in utilization levels as the risk averseness level increases. For this reason, each OR manager can identify with a particular averseness range that suits their surgical case environment.

Similarly, we also showed how encouraging open posting can affect the number of ORs used. We observed that shifting from using no open posting encouragement (enforcing a penalty of e = 0.3) to high open posting encouragement (i.e., no penalty or e = 0.0) decreased the number of ORs scheduled by 15%. This implies that for every 33% increase in open posting encouragement issued, there is one fewer OR required in the schedule. The results also emphasize the impact of open posting rooms and how important scheduling accuracy is to running an efficient OR. These models provide managers options when deciding on a new OR schedule and should give managers further motivation to seek more accurate case posting times.

In the SPPM, we focus on assigning surgeries to rooms, but not necessarily demanding a specific sequential order for the cases. Equipment limitations and block release times were also not considered, as the primary focus was to identify the change in rooms used and utilization based on the manager's risk tolerance and willingness to allow open posting. These are items that we are incorporating into our future research. One can infer that the results may be better than what can be achieved in practice. However, the management team often makes many day-of-surgery adjustments to improve performance that are difficult to capture in a model. Thus, the research outcomes likely still provide a reasonable estimate of OR performance.

Concerning block release, one key outcome of this research is that there would be less need for OR managers to make day-of-surgery adjustments if blocks were set and then not adjusted or released. Yet, we still see the need for future research that involves a more dynamic approach to the posting of cases and the events leading up to the day-of-surgery outcomes. This will provide insight into how and when cases are posted, and when (if any) block release time should be set. Separate from the modeling framework, there is great value in improving the case duration or schedule accuracy. Even slight improvements in booking times matching actual case duration could make dramatic changes to efficiency levels observed.

CHAPTER III
OPTIMIZING HOSPITAL BLOCK RELEASE POLICIES THROUGH THE USE OF SIMULATION

1. INTRODUCTION

Due to different surgery types and characteristics, certain surgical groups may be able to plan their surgery schedule farther ahead than others. It is a common practice at many hospitals to allow surgical time to be “released” when the allocated block time will not be needed or used. This allows for excess and add-on cases to be added to the surgery schedule.

The policy set forth by the hospital for the release time can affect how and when surgeries are scheduled. If a surgical group holds on to their time for too long and does not fill their block, they will be penalized for having a low room utilization. However, if a surgical group releases their unfilled time, they will only be accountable for what they have scheduled.

In this paper, we use simulation techniques to find suggested block release times for particular surgical groups using the ideas of blocked and open posting time to create the Master Surgical Schedule. We provide results and discussion about these policies and how scheduling accuracy and variability may impact these decisions.

2. LITERATURE REVIEW

Block release times and methods to obtain optimal policies have not been studied much in the literature.

Dexter et al. [30] present a "descriptive study" of case scheduling, add-ons, and cancellations a week before the day of surgery. They found that surgical groups that had filled their blocks and needed more surgery time did this about two days prior to the day of surgery. They also noted that at least half of the ORs studied had their last case scheduled or changed within two days before surgery. They concluded that, "There are so many changes made to so many cases in so many ORs even 2 workdays ahead, that making plans more than 2 workdays ahead is unlikely to be productive." May et al. [31] found similar results in their literature review of current surgical scheduling techniques. They pointed out the high frequency of changes to the surgical schedule.

Dexter and Macario [32] analyzed when block time should be released by adding hypothetical cases to ORs with excess allocated time. They studied cases scheduled on the day of surgery, morning of the day before surgery (shortest possible), three days before the day of surgery, and five days before the day of surgery. Cases of length one, two, and three hours were tested to represent actual cases. They concluded that, most often, adding new cases prior to the day of surgery reduced overutilized time better than releasing block time three to five days before surgery. This showed that OR time release has a negligible effect on OR efficiency. Dexter and Traub [18] also noted that there is no advantage in terms of OR efficiency to releasing a service's allocated OR time until there is a case to be scheduled into that OR time. They explain that: "A few days before surgery, maximizing the efficiency of use of OR time is synonymous with minimizing overutilized OR time."

However, despite these results, medical doctors Dr. Mazzei and Dr. Blasco [33] feel that "Variable release times are a quick way to build flexibility into a schedule." Their suggested list of block release times ranges from one day for Burn service and Cardiac specialties to 14 days for Orthopedics (joint) and Plastic (cosmetic).

Other papers have investigated the role of block release time when scheduling new cases into the block. Dexter et al. [34] examine impacts on OR utilization levels when deciding which surgical group should release their block time. They concluded that scheduling new cases into rooms that have the largest difference between scheduled and allocated time, rather than the room with the most unscheduled time, has a better effect on OR efficiency levels.

Discrete-event simulations in the field of healthcare have begun to increase over the years according to Jun et al. [35] and Hamrock et al. [36]. Both papers deem simulation as a cost-effective tool to help allocate resources, improve patient flow, manage bed capacity, schedule staff and surgeries, and more. An example of this work can be seen from Dexter et al. [37], who conducted a computer simulation to analyze methods affecting utilization and efficiency. They concluded that OR utilization was mainly affected by how far in advance a case was posted. Cases requested further in advance are better able to be fit into remaining OR time in order to better fill the block. Additionally, M. Persson and J. Persson [38] use discrete-event modeling to simulate two management policies and their effects on patient wait time, number of surgeries cancelled, and OR utilization. They also consider cost in reference to the amount of overtime incurred. Probability distributions based on historical data were used to generate patient arrivals and case durations.

3. METHODS, ASSUMPTION, AND APPROACH

A two-room system, one blocked and one open posting, was created and tested in our simulation model. Figure 2.1 shows the flow of the simulation relating to how and when a surgery is booked in an OR. Once a surgery is called in and assigned relevant information (Modules 1A,B and 2A,B), it is immediately placed in its proper OR if there is time available on the schedule (5A,B). If there is no time available and the surgery was posted to the blocked room, the surgery will immediately check if there is time available in the open posting room (7,8A). If there is no time in either room, the case is recorded as not posted to the schedule (9A,10B). Rooms cannot be overbooked. If there is no time available and the surgery posted to the open room, it will go into a holding queue, where it will wait for block time to be released (7B). Once the block is released, the surgery may check the blocked room to see if there is available time for it to occur (8B). If block release is occurring, cases will be sent to the room with the largest amount of available space, or unfilled block time, compared to its total allocated time. Dexter et al. [34] found that this was a good way to maximize efficiency levels. Number of cases successfully entered in another room, number of cases not posted to the schedule, and room utilization levels are tracked. Different arrival patterns, flows, and block release times are tested to suggest the best release policies based on the arrival distribution and expected room utilization.

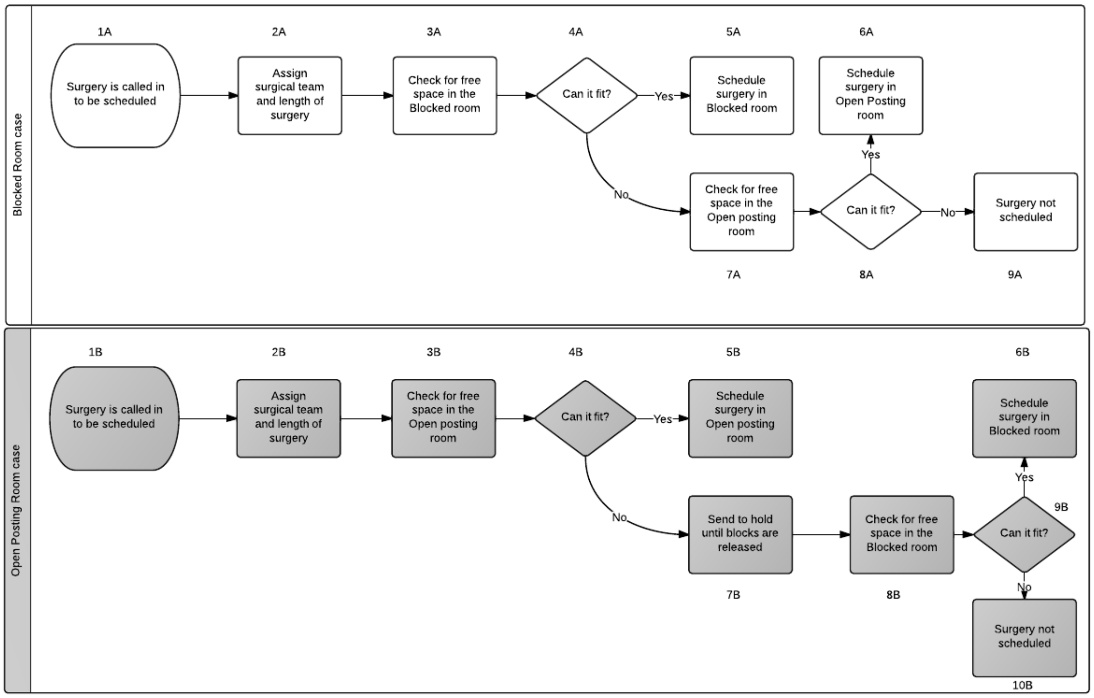

Figure 2.1: Simulation flows for cases arriving to the blocked and open posting rooms

The cases that are scheduled in each room are then played out. For purposes of simplicity with this initial testing, we let the scheduled surgery duration also act as the actual surgery duration. These durations can later be manipulated to allow for more variation in actual duration lengths. Once the surgeries play out, the utilization of each room is calculated by taking the total sum of the surgery durations divided by the room length.

Room lengths were set to be 10 hours, each with a simulated caseload of 60%, 80% and 100% of room capacity, on average. In other words, the surgery arrivals (or caseload) provided in the three scenarios would result in an expected utilization of 60%, 80%, or 100%, respectively. These utilizations provide adjustments for hospitals during slow or peak times while also showing the equivalent of what might be an 8, 10, or 12-hour room respectively.

The surgery arrivals for each room type, shown in Figure 2.2, were kept separate in order to test different flows and number of cases. The arrival patterns, referred to as “Early,” “Middle,” and “Late,” peak seven, five, and three days prior to the day of surgery, respectively. The distributions all have the same peak number of surgery arrivals and uniformly increase until they hit the peak, then decrease until they hit zero or do not have any days left. The “Early” and “Late” distributions are mirrors of each other, and all arrival patterns have the same number of cases.

These patterns, verified and generalized by data from the Medical University of South Carolina (MUSC), were used to encompass a wide variety of specialties and to imitate arrival patterns of surgical groups who plan close to and far ahead of the day of surgery. We take a ten-day span into consideration for our initial analysis because it allows for at least a week of posting cases and up to three days of block release time. These distributions can be adjusted to fit other representative or specific case posting patterns.

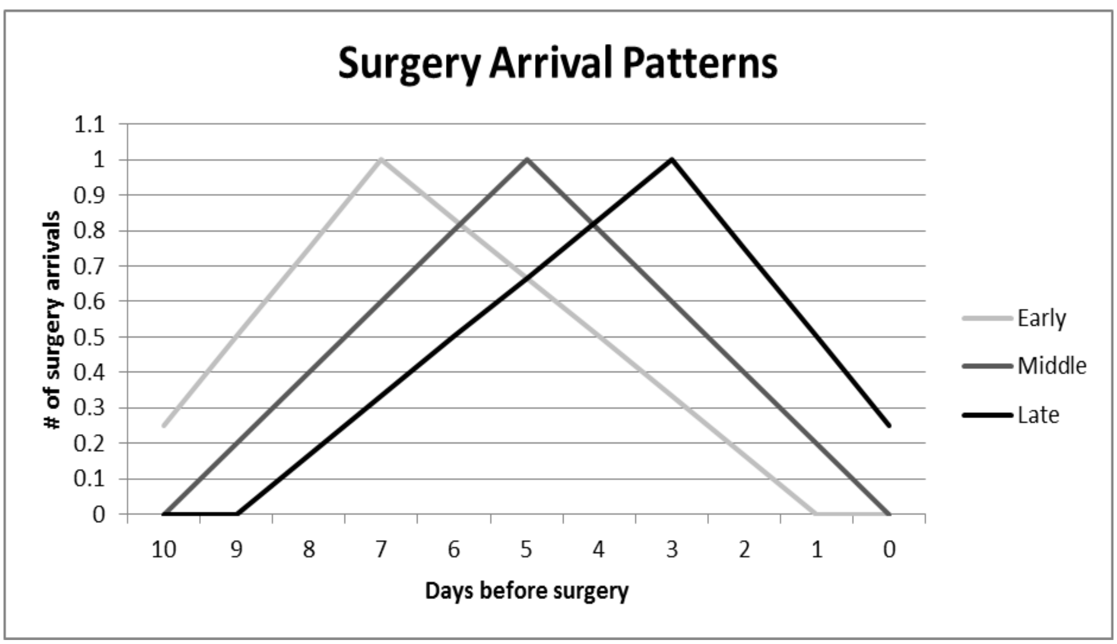

Figure 2.2: Base case scenario surgery arrival distribution patterns with three different peak times

In addition to testing different arrival patterns, we used four duration scenarios (as shown in Table 2.1). The Short, Medium, and Long surgical durations represent a spread of the types of surgical cases services encounter. The Base Surgical duration, derived from the duration used in Dexter and Macario [32], allows for more variability in case lengths. The other three durations are slight variations of the Base case. The expected number of surgery arrivals (or postings per room) was calculated by taking the room length (10 hours) and dividing it by the average length of the duration for each scenario.

Table 2.1: Surgical duration scenarios

Code	**Description**	**# of Case Arrivals**
Base Surgical Duration	Uniformly distributed between 1 and 3 hours	5
Short Surgical Duration	Uniformly distributed between 1 and 2 hours	6.67
Medium Surgical Duration	Uniformly distributed between 1.5 and 2.5 hours	5
Long Surgical Duration	Uniformly distributed between 2 and 3 hours	4

4. ANALYSIS

A total of 432 simulations were run using Arena's Process Analyzer with 100 repetitions each. For each of the four surgical duration scenarios, nine surgery arrival pattern combinations, four block release times (0, 1, 2, and 3-day release), and three expected utilization levels (60%, 80%, and 100%) were tested. Rooms were allowed to have different arrival distributions, but surgical durations and expected utilizations remained the same. Results from the different arrival combinations were combined and averaged. Room utilization levels and percent of cases that could not be added were the main statistics used to suggest the best block release time from these simulations.

An initial test was run on the Base case surgical duration using block releases ranging from no release to a 5-day release. We define release times as the number of days prior to surgery that excess or unfilled surgery time in a blocked room becomes available to other surgical groups. When the block is released, any cases that could not fit in the open posting room can have a chance at being scheduled in the blocked room's excess time. As shown in Figure 2.3, there is a 1-2% increase in combined room utilizations when switching from no release to having release. There is very little difference in average room utilization across all of the scenarios with block release (<1% increase or decrease).

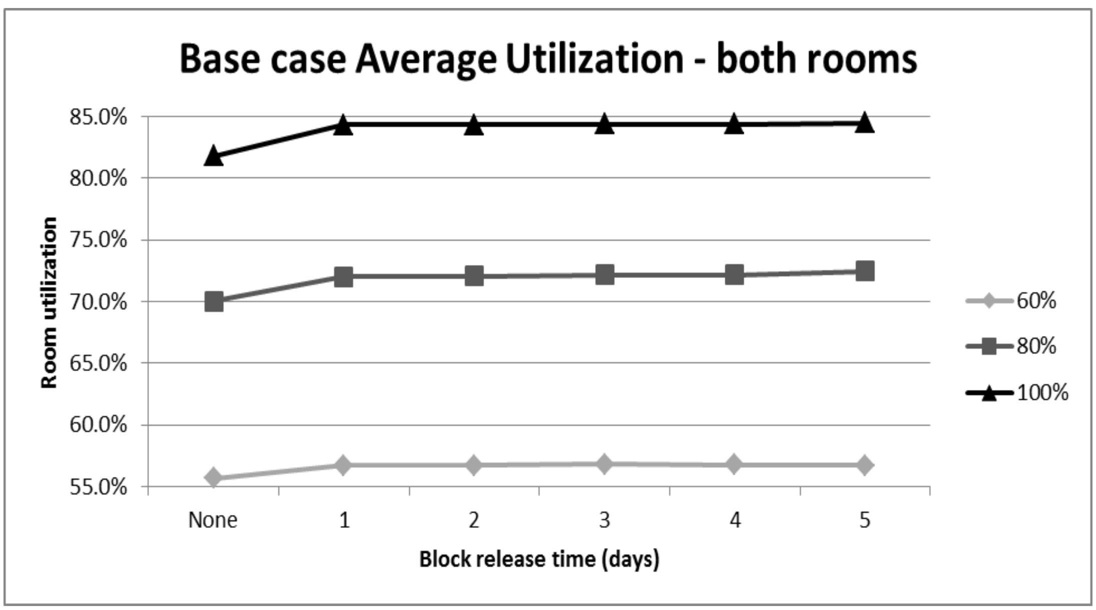

Figure 2.3: Effects of six block release policies on average room utilization levels

Similar results are shown in Figure 2.4, which displays the percentage of cases not posted to the schedule for the same release periods. Once again we see a 1-2% decrease in average unscheduled cases when switching from no release to having block release, while there is very little variation across the individual scenarios with block release.

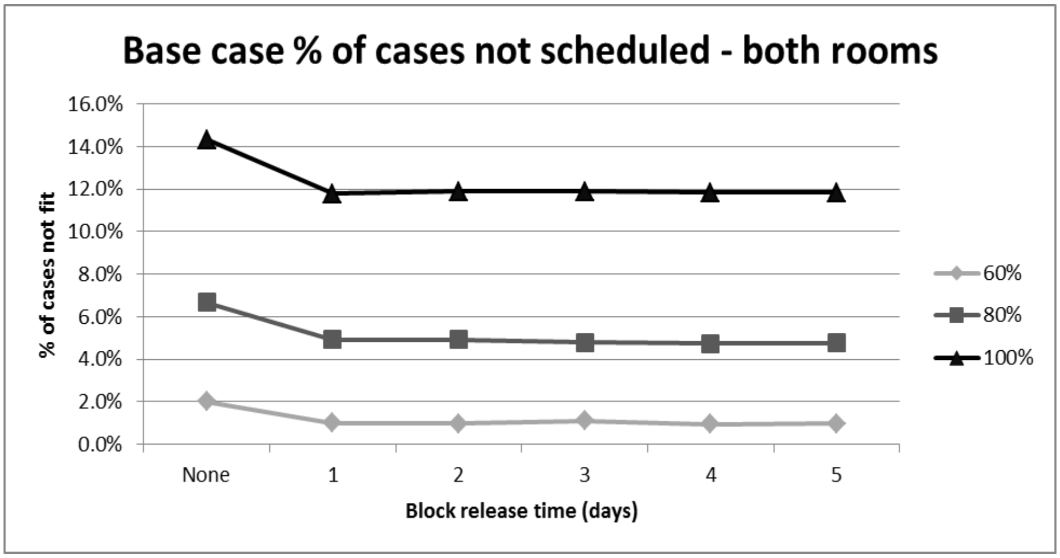

Figure 2.4: Effects of six block release policies on unscheduled cases for both rooms

Because there was very little difference across the scenarios with block release, long block releases are not frequently used, and our sponsoring hospital specifically had an interest in a 3-day release, comparing no release against a 3-day block release seemed most appropriate. Results from each test were examined for each room specifically, as well as both rooms combined.

The results showed that using a 3-day block release led to higher room utilization and a lower percent of cases not scheduled for the blocked room, while no release was preferred for the open posting room. Figure 2.5 shows a comparison of these release policies and their effect on room utilization levels for the combined room utilization of the blocked and open posting rooms for every scenario. Overall, the three-day block release policy showed slightly higher room utilization levels than a no release policy. A block release policy allows for more cases to be (or tried to be) fit into the blocked room if there is available space, which boosts the room utilization of the blocked room. Without a block release policy, cases that do not fit into the open posting room are sent away, even if they were able to fit into the blocked room.

While it seems unrealistic that cases that could have fit into a block do not get scheduled, it is indicative of the overall picture of not having forced block release. If a surgical group holds on to their block time, despite not having enough cases to completely fill the block, their calculated room utilization is often punished because they are being judged on using their entire block, rather than a subset of their allocated time. A similar punishment occurs here. By not allowing extra cases to use the extra time, the service receives an overall lower room utilization rate.

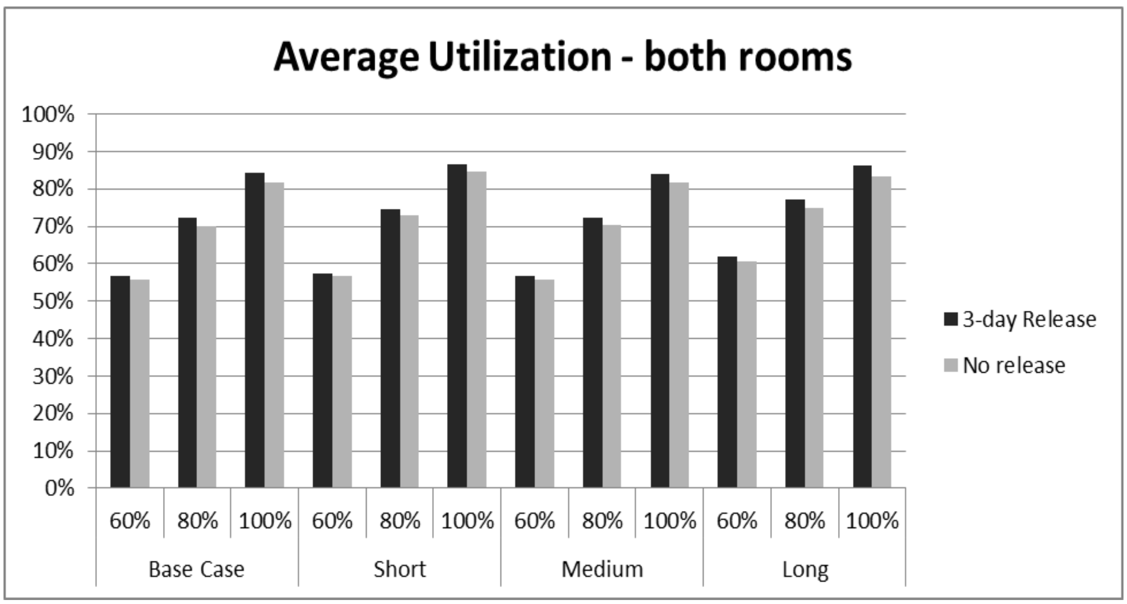

Figure 2.5: Effects of two block release policies on average room utilization levels

Figure 2.6 shows the differences in utilization when the two rooms are examined separately. To clarify, the difference takes the room utilization received from a 3-day release less the no release policy. The positive differences indicate higher utilizations occurred in the blocked room, while negative differences indicate higher utilizations occurred in the open posting room.

When examining the blocked room, we see an 8-11% difference (in favor of the 3-day release) in room utilization levels between the two release scenarios. For the open posting room, we see a 2-9% difference (in favor of no release). Due to the magnitude of each difference, with a larger difference for the 3-day release, when we view the combined utilizations, there is only a 1-3% difference (in favor of the 3-day release). This poses some interesting insights about the pros and cons of protecting a service's block, which will be discussed in further detail in the Managerial Insights section.

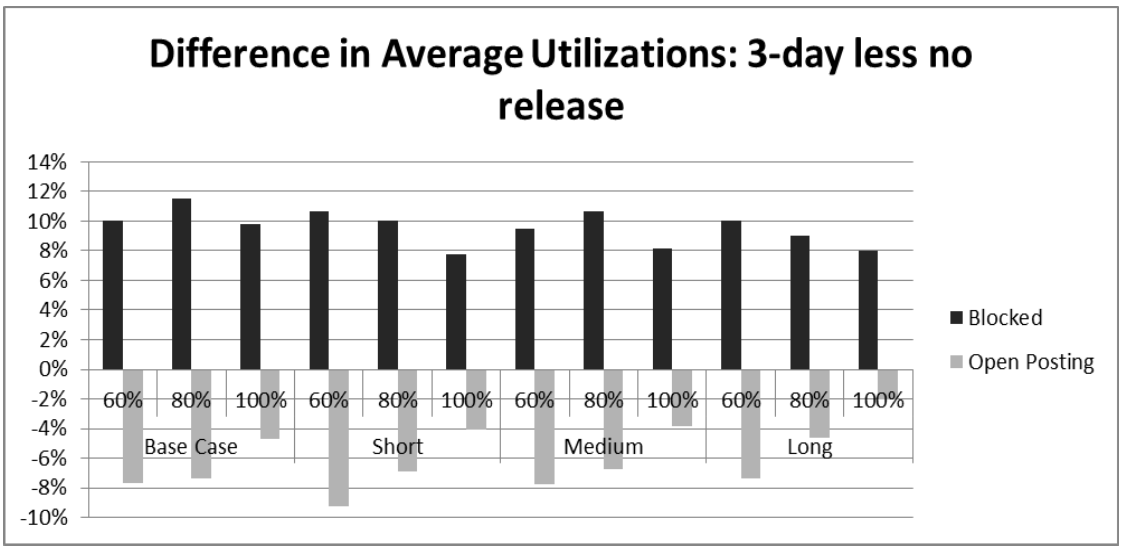

Figure 2.6: Blocked / Open Posting room utilization differences from 3-day release to no release

In addition to tracking room utilization levels, the number of cases not scheduled was also recorded. This statistic is meant to represent cases that are either delayed and set for another day or for cases that will find another hospital completely. According to Macario et al.[39], this is one of eight strong measures of OR efficiency. Figure 2.7 shows a comparison of the combined percentage of these numbers for the blocked and open posting rooms. Once again, the results favor the 3-day release policy since the no release policy does not allow for open posting cases to be sent to the blocked room. This leads to more cases not being posted to the schedule.

As expected, the distribution arrival pattern impacted the percentage of cases that could not be scheduled. For the Base Case and Short scenarios (uniformly distributed between 1 and 3 hours and 1 and 2 hours respectively), shorter case durations allowed for a better and easier fitting of cases while the Long scenario (uniformly distributed between 2 and 3 hours) rooms were more difficult to fill completely.

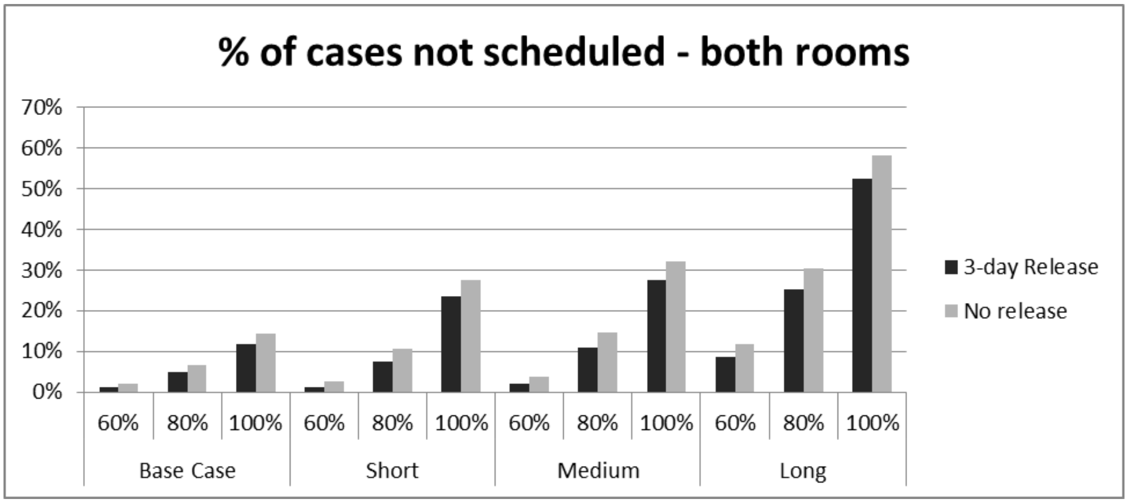

Figure 2.7: Effects of two block release policies on unscheduled cases for both rooms

Figures 2.8 and 2.9 show the effects on the percentage of cases not scheduled for the individual rooms. When the rooms are examined separately, there’s a much more significant difference than when averaged together. There is a 1-10% difference (in favor of the 3-day release) in percent of cases not scheduled for the blocked room and a 3-15% difference (in favor of no release) for the open posting room. The percentage of total cases, across both rooms, only shows 1-2% difference (in favor of no release).

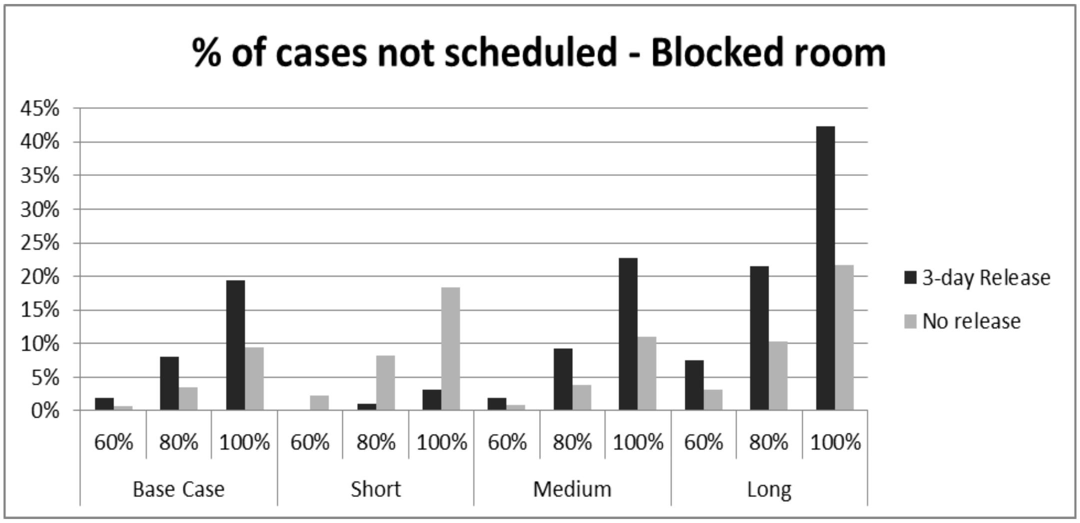

Figure 2.8: Effects of two block release policies on unscheduled cases for the blocked room

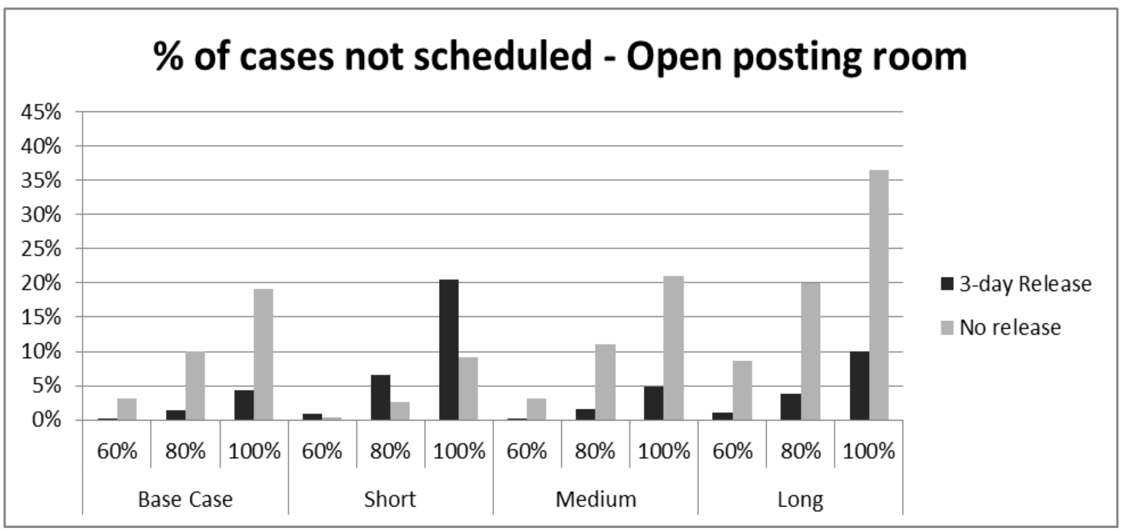

Figure 2.9: Effects of two block release policies on unscheduled cases for the open posting room

5. RESULTS

For the given scenarios, the 3-day block release favored higher room utilizations and a lower percentage of unscheduled cases for the block-specific room, while no release favored the open posting room. While our results show a preference for the 3-day block release, upon close examination of the data, we can gain knowledge about the benefits of protecting a surgical group's block.

For services that are active and highly utilizing their blocks (at or above 80% expected utilization levels), forcing the blocked room to release time increases the number of cases not booked by about 5% and increases utilization by about 9% (Figure 2.5). For services that are already operating at a high utilization level, the increase is likely not needed and is not a worthwhile tradeoff. On the other hand, if the service is operating below the 80% mark, forcing a 3-day release might be beneficial to raise the utilization without losing as many cases. On average, a surgical group with a 60% expected utilization will only lose about 1.25% of their cases for that day and gain about a 10% increase in utilization.

We can also look at this information by average case duration length. Blocked rooms that are paired with surgical groups that have shorter, faster cases may expect an 8% decrease in the number of cases not booked and a 9% increase in utilization with a 3-day block release. When paired with surgical groups with longer case durations, blocks may experience a 12% increase in the number of cases not booked but a 9% increase in utilization with forced block release. When paired with surgical groups with medium duration cases, results lie somewhere in the middle, with an expected loss of about 6% of cases and an expected increase of 10% utilization.

Table 2.2 shows the average result in changes of room utilization and percentage of cases not booked when switching from no block release to 3-day block release <u>*for the blocked room only*</u>. In every case, switching from no block release to having block release allows for an increase in utilization. However, forcing a block release is likely only necessary when the block is running below the 80% expected utilization mark because the loss in cases does not balance out the utilization increase, except for groups paired with shorter surgical durations.

Table 2.2: Differences when switching from no release to a 3-day release for the blocked room

	60%		**80%**		**100%**	
	Utilization	% of cases not booked	Utilization	% of cases not booked	Utilization	% of cases not booked
Base	+10%	+1%	+11%	+4%	+10%	+10%
Short	+10%	-2%	+10%	-7%	+8%	-15%
Medium	+10%	+1%	+11%	+5%	+8%	+12%
Long	+10%	+5%	+9%	+11%	+8%	+20%

Table 2.3 shows the average result in changes of room utilization and percentage of cases not booked when switching from no block release to 3-day block release for both rooms. In every case, switching from no block release to having block release allows for an increase in room utilization and a decrease in unscheduled cases. Unlike the results in Table 2.2, because the blocked room receives a large increase (~10%) in room utilization and the open posting room receives a decrease (~7%) in room utilization, when averaged together the combined utilization does not nearly change as much. We also note that the 3-day release policy's decrease in the number of unscheduled cases means that some cases from the blocked room must occur in the open posting room.

Table 2.3: Differences when switching from no release to a 3-day block release for both rooms

	60%		**80%**		**100%**	
	Utilization	% of cases not booked	Utilization	% of cases not booked	Utilization	% of cases not booked
Base	+1%	-1%	+2%	-2%	+3%	-2%
Short	+1%	-1%	+2%	-3%	+2%	-4%
Medium	+1%	-2%	+2%	-4%	+2%	-4%
Long	+1%	-3%	+2%	-5%	+3%	-6%

A final comparison test was conducted in order to identify differences between arrival pattern combinations. Nine combinations of the patterns shown in Figure 2.2 were used for the previous tests, but their results were combined and averaged together. Figure 2.10 shows the results from each of the nine combination patterns using the Base case surgical duration and 80% expected utilization levels. The highest utilization for both release policies occurs when the blocked room has Late arrivals, or a peak occurring on day 7 out of 10, and the open posting room has Early arrivals, or a peak occurring on day 3 out of 10. The worst utilization for both release policies occurs when the blocked room and the open posting room both have Early arrivals.

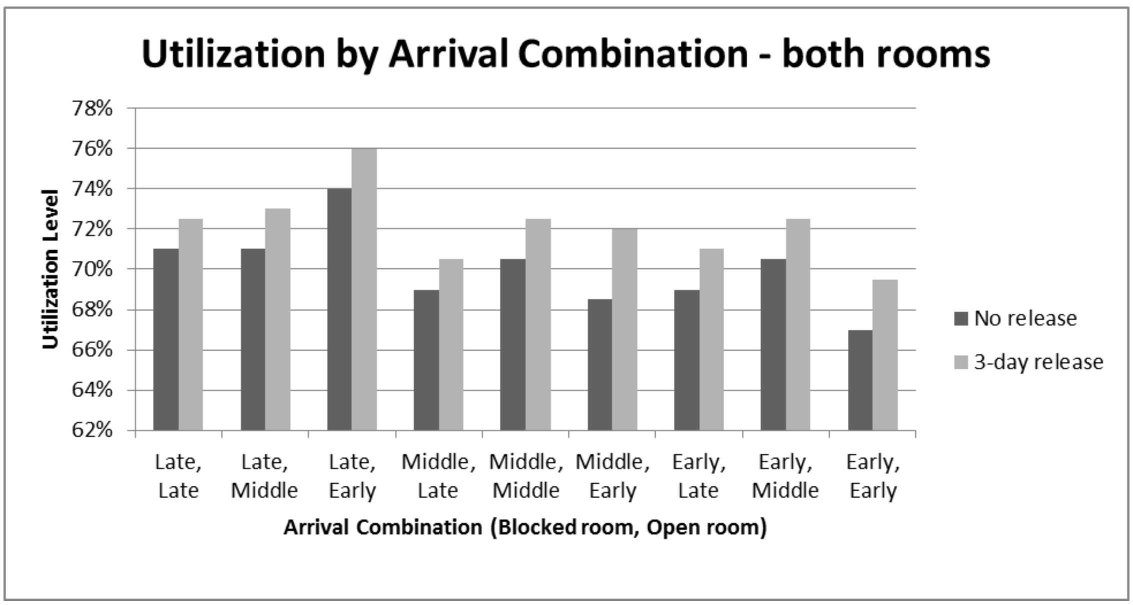

Figure 2.10: Utilization levels for the Base case with expected utilization levels of 80%

On the other hand, for the number of unscheduled cases, the results are not the same. The combination with the fewest number of unscheduled cases occurs when the blocked room receives Middle arrivals, or a peak occurring on day 5 out of 10, and the open posting room receives Late arrivals, or a peak occurring on day 7 out of 10. The most unscheduled cases occur when both rooms receive Early arrivals, or peaks occurring on day 3 out of 10. Overall it seems that a combination of Late arrivals with either a Middle or Early arrival pattern will give positive results.

These results may be due being able to make better scheduling decisions and fit cases together when blocked cases arrive later, rather than immediately filling up the block when they arrive sooner. This is because if there is block release occurring, cases will be sent to the room with more available space, rather than the room to which they were assigned. If no release is occurring, late-arriving cases may not make it to the schedule in time for the day of surgery.

6. MANAGERIAL INSIGHTS

According to MUSC's OR Analytics Manager Charles Hajzus, significant differences in surgery posting distributions arise from natural and self-inflicted variation. Self-inflicted variation may occur due to ineffective schedulers, staff or policy-related issues, inefficient clinic setups, equipment issues, and backlog. However, the focus should be on reducing the natural variation because it is easier to defend and will last over time.

We provide a variety of surgery arrival patterns to encompass the variation and prioritization between surgical groups and their cases. When viewed individually, release policies may favor one type of room over another. Our results are intuitive with how cases are posted to each room – the 3-day block release allows for better utilization of the block-specific room because it is allowed to take on more add-on cases, while the open posting room will continue to accept all cases for which it has space.

Our overall results match the suggestions of Dexter and Macario [31], whose study showed that block release has a negligible effect on OR efficiency when rooms receive properly allocated time. However, our results showed that individual rooms can have significant impacts on utilization and unscheduled cases that when looked at as a whole, are often minimal. These utilization changes may appear to be minimal when averaged together due to utilization increases offsetting the decreases.

A standard, rather than optional, block release may have positive influences on surgical groups that do not release their time in hopes of filling it with an incoming case. With a forced block release time, services can be more confident about getting a last minute case scheduled while keeping high utilization levels and low unscheduled case rates for their specific OR.

7. CONCLUSIONS AND FUTURE WORK

This paper presented the results of two main block release policies and showed their effects individually and combined on a small-scale simulation of two ORs.

One thing to consider when reviewing these results is how the setup of the block schedule can affect the need (or lack thereof) to have block release. Epstein and Dexter [26] found that a utilization level of 80% was a desirable rate for many hospitals. A block schedule that is set up with assigned blocks that allow for an 80% utilization will not necessarily need block release. However, if services are performing at the 60% or below rate, it might be more beneficial to adjust the block size to better fit the service’s needs, rather than forcing them to have block release. McIntosh et al. [2] emphasize the importance of allocating proper block sizes two to three months prior to the day of surgery in order to increase OR efficiency. Dexter and Macario [31] also emphasize this point, suggesting that block release can be decided by the political environment as long as proper time allocation occurs.

The work presented is a small piece of a larger, untapped field of OR and block release research. Future work will consider different surgical distributions and expected utilization levels for each room. Longer spans and block release periods, smarter surgical duration lengths, and accuracy of scheduled versus actual surgery times and their impacts on release policies may also be items to investigate.

CHAPTER IV
OPTIMIZING HOSPITAL BLOCK RELEASE POLICIES THROUGH THE USE OF SIMULATION: A CASE STUDY

1. INTRODUCTION

Discrete-event simulation (DES) is becoming a popular tool as the healthcare industry undergoes more pressure to increase efficiency and provide better care at lower costs [40]. The use of simulation allows for management policies to be tested, prior to implementation, to see the effects on individual processes and the system as a whole [36].

In this chapter, we will simulate a case study to test policies involving the allocation of OR time to surgical groups and the release of this time for other groups to use. The work will be done in two steps: First, time will be allocated using the mathematical models in Chapter II. Three schedules will be created for each data set and location to show differences in the policies chosen. Second, a more advanced simulation model, built from the preliminary work discussed in Chapter III, will be used to test the schedules created with different block release policies. Two release policies will be examined in conjunction with the schedules created in order to make observations and suggestions based on schedule type and surgical group characteristics.

2. LITERATURE REVIEW

Jun et al. [35] provide a compilation of healthcare studies that have used DES in areas such as patient scheduling, flow, and availability of resources. They noted that detailed simulation models can be a follow-up to optimization outputs in order to identify and analyze performance measures and gain acceptability.

Davies and Davies [40] found that simulation modeling was well-suited for healthcare clinics due to its ability to capture complexities of medical systems.

Steins et al. [41] used DES to identify bottlenecks in post-operative care in order to more evenly distribute utilization levels across ORs. They tested four scenarios and analyzed throughput, resource utilization, and postoperative resource numbers.

Similar to our work, Persson and Persson [38] use DES and optimization modeling to study the impacts of various patient arrivals, surgical durations, and OR allocation times. They focus on three main processes: patient arrival, patient queuing, and OR scheduling. Patient arrivals are modeled by a Poisson process based on one year of historical data of orthopedic surgeries. Expected surgical durations are taken from a mean estimate based on the surgery type, while actual durations are modeled according to a lognormal distribution.

The concept of accuracy in surgical scheduling is important to reduce under or overutilization of the OR. "Estimation errors not only affect the flow of patients through operating theatres, but the coordination of activities in such theatres and in the care giving process" [42].

Zhou et al. [43] found that solely using historical data to estimate surgery times is not effective. Dexter et al. [44,45] noted that the procedure type was the most important factor in predicting case duration. Combes et al. [46] noted the importance of using a patient's profile to evaluate surgery durations and suggested a process in knowledge discovery of databases (KDD) to estimate these durations. Strum et al. [47] found that the surgeon and type of anesthesia received can help predict variability and deviation lengths.

3. METHODS, ASSUMPTION, AND APPROACH

A larger system was built from the simulation model used in chapter II. Three sets of historical data were obtained from the same level one trauma, teaching hospital in the Southeastern United States of America in Chapter I. The data was used to obtain information from 24 services at 2 locations. For this work, the third location was left off due to its OR time constraints and specific surgery types.

One particular day of the week's data was used to create the Master OR schedule, simulate surgery arrivals, and have the surgeries occur. For consistency, this day of the week is the same as the one shown in the graphs in Chapter I.

For each location and data set, twelve months of historical data were used to set the Master OR schedule using undertime-to-overtime ratios (*h*) of low (*1.25*), medium (*1.75*), and high (*2.25*) averseness to overtime levels, as shown in Table 1.1. A constant encouragement level (*e)* of Medium-Low was used, as shown in Table 1.2. A total of 13 ORs for low averseness to OT levels and 14 ORs for medium and high averseness to OT levels were used in Location 1, while seven ORs for all averseness to OT levels were used in Location 2. Using the next three months of historical data, distributions were created in Arena's Input Analyzer. Surgery arrival, case booking length, and accuracy level distributions for each group that was assigned a block of time. Groups that did not receive a block of time were combined to create the Open Posting group, which then received the same distributions. All distributions were ranked on lowest squared error and had p-values greater than .05 when possible.

The simulation model is similar to the one found in chapter II. Cases arrive and are noted as being from their particular surgical group. Based on the assigned booking time, cases check to see if there is free space in their assigned operating room, adding a standard 30 minutes for turnover time for all cases but the last. If a case from a blocked room cannot fit, it will immediately check to see if there is space in the open posting room. Open posting cases, on the other hand, check to see if there is space in the open posting room first. If there is not space, they will go into a hold block until blocks are released. Once blocks are released, these cases will be assigned a new room to see if there is space available. New cases that arrive after blocks have been released will go to either their blocked room or the open posting room, whichever has more unfilled time.

In order to play out the surgeries and record utilization levels, an accuracy distribution was created to identify and acknowledge differences between scheduled and actual surgery durations. Accuracy levels were taken from the historical data by dividing the actual surgery duration by the booking duration. An accuracy level of 1 indicate the surgery was perfectly on time, accuracy levels above 1 indicate the surgery ran late, and accuracy levels below 1 indicate the surgery ended early. While ORs were filled up with the scheduled durations, the utilization results shown in the next section are from the actual durations in the rooms. Additionally, the number of unscheduled cases was also recorded and is shown below.

4. RESULTS

A total of 147 simulations, 93 for Location 1 and 54 for Location 2, were run using Arena's Process Analyzer with 100 repetitions each. Similar to the tests run in Chapter II, we define release times as the number of days prior to surgery that excess or unfilled surgery time in a blocked room becomes available to other surgical groups. When the block is released, any cases that could not fit in the open posting room can have a chance at being scheduled in the blocked room's excess time.

We classify each surgical group that receives block time in terms of surgery arrivals, case durations, and accuracy levels as shown in Tables 3.1 and 3.2 for locations 1 and 2 respectively. Groups that have average arrivals between 1 to 10 days are considered Early, 11 to 20 are considered Medium, and 20 to 30 are considered Late. Groups whose case durations average between 0 to 100 minutes are considered Short, case durations between 101 to 200 minutes are considered Medium, and case durations between 201 to 300 minutes are considered Long. Accuracy levels that range between 1.0 to 1.12 are considered Strong, levels between 1.13 and 1.24 are considered Medium, and levels between 1.25 and 1.36 are considered Weak.

Table 3.1: Classifications of surgical groups that received block time at location 1.

	Arrivals	**Duration**	**Accuracy**
8	Medium	Medium	Weak
11	Early	Medium	Medium
12	Medium	Medium	Medium
13	Late	Long	Medium
18	Medium	Medium	Medium
22	Early	Medium	Strong
23	Medium	Medium	Weak
7	Medium	Medium	Medium
9	Medium	Medium	Medium

Table 3.2: Classifications of surgical groups that received block time at location 2.

	Arrivals	**Duration**	**Accuracy**
4	Early	Long	Medium
7	Medium	Medium	Strong
19	Late	Medium	Weak
21	Early	Long	Strong
24	Early	Medium	Medium

Each surgical group received their own distribution for each classification and received blocked hours based on the OT-to-UT ratio used for the schedule. Figures 3.1 and 3.2 show the average utilizations for each surgical group using 3-day release and no release policies at location 1. There are not noticeable utilization differences between risk ratio scenarios, but there are differences between release policies. Switching from a 3-day release policy to a no release policy increased utilization by an average of 10% across all sets. The largest utilization increases were seen by groups 12 and 18, while the smallest utilization increases were seen by groups 13 and 22.

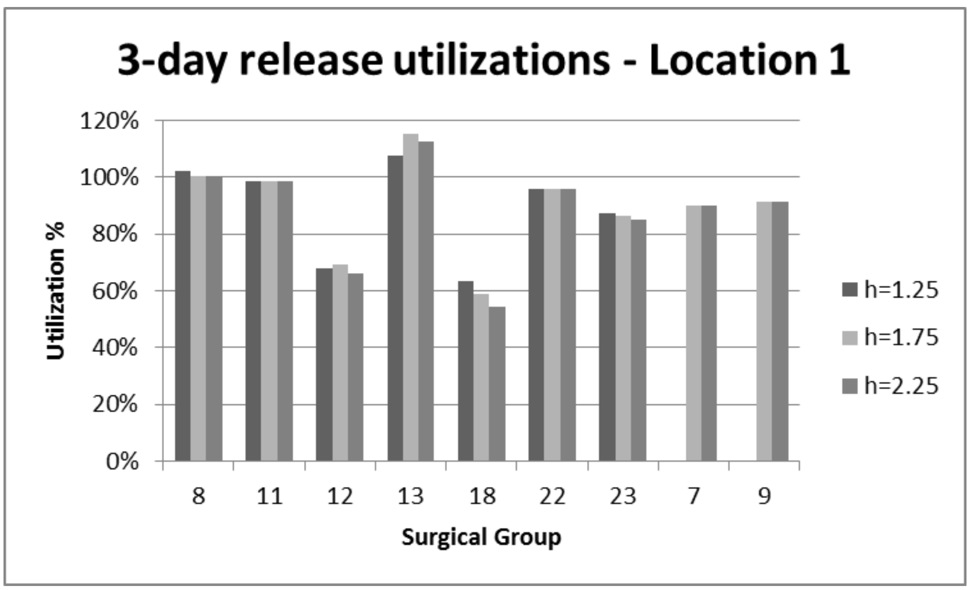

Figure 3.1: Average utilizations across sets of the 3-day block release policy for three risk ratios

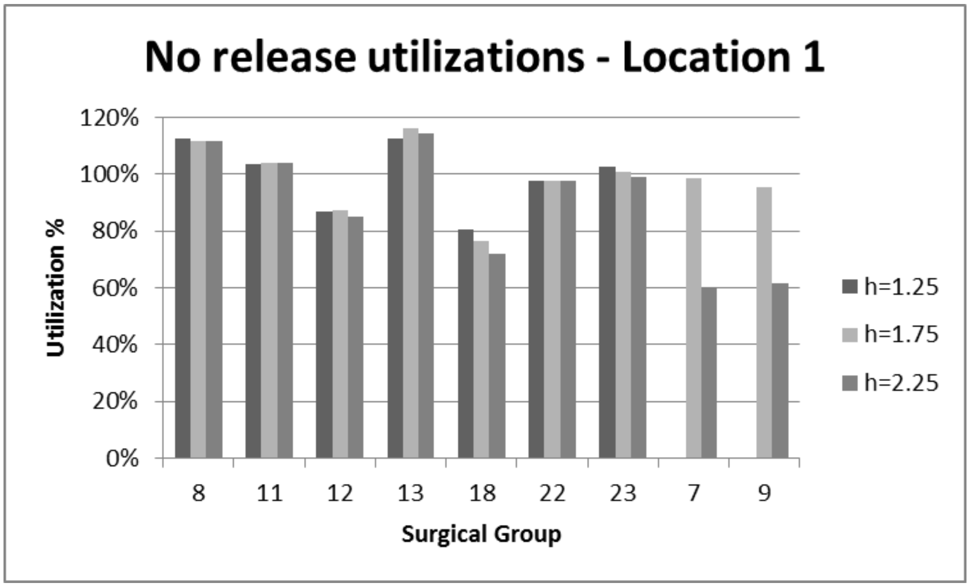

Figure 3.2: Average utilizations across sets of the no release block policy for three risk ratios

Figures 3.3 and 3.4 show the average utilizations for each surgical group using 3-day release and no release policies at location 2. Once again, there are not noticeable utilization differences between risk ratio scenarios, but there are differences between release policies. Switching from a 3-day release policy to a no release policy increased utilization by an average of 4% across all sets. The largest utilization increases was seen by group 19, while the smallest utilization increases were seen by groups 4 and 7.

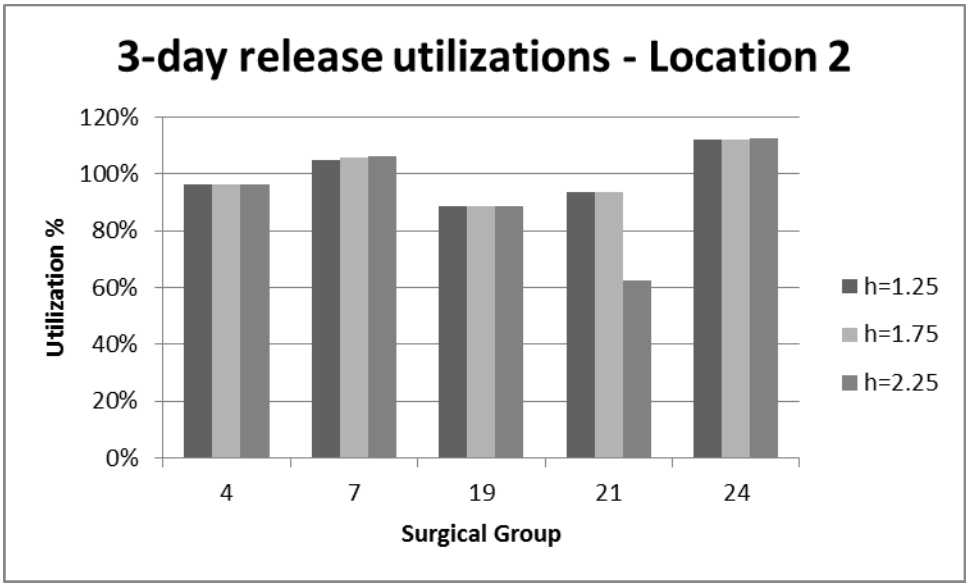

Figure 3.3: Average utilizations across sets of the 3-day release block policy for three risk ratios

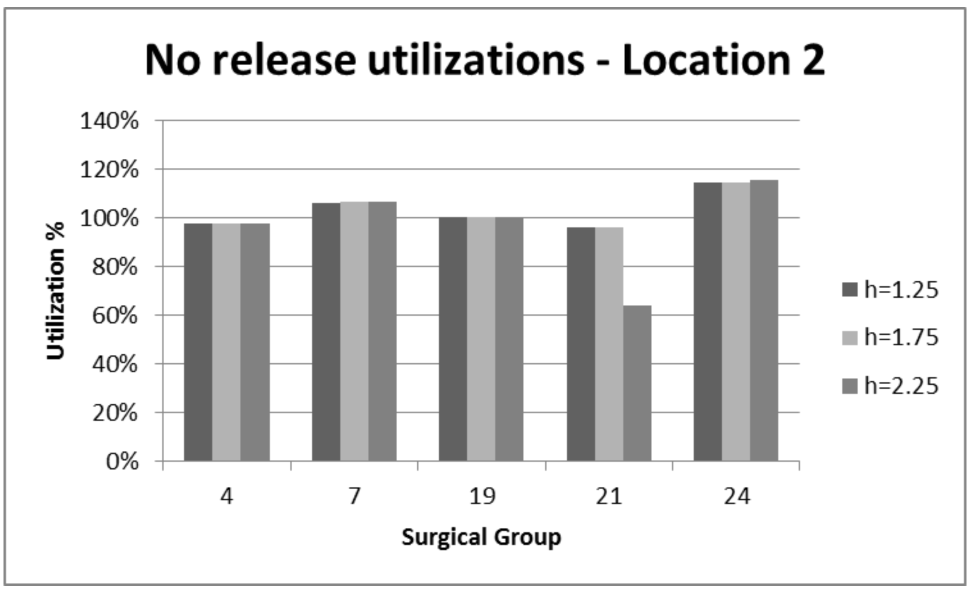

Figure 3.4: Average utilizations across sets of the no release block policy for three risk ratios

Both locations show very high utilizations for their groups, which may be due to the differences in accuracy levels between booking and actual surgical durations. To address this potential issue, we ran simulations for the Medium-Low averseness to OT case (h=1.75) at location 2 but used an expected utilization to help buffer the differences in durations. Figure 3.5 shows the average utilizations for the three expected utilizations in Chapter II. The expected utilization provides a buffer that does not take on certain additional cases because it knows extra time will be used for the current cases. As the results show, the utilization levels for each group are closer and more reasonable to what a hospital would desire.

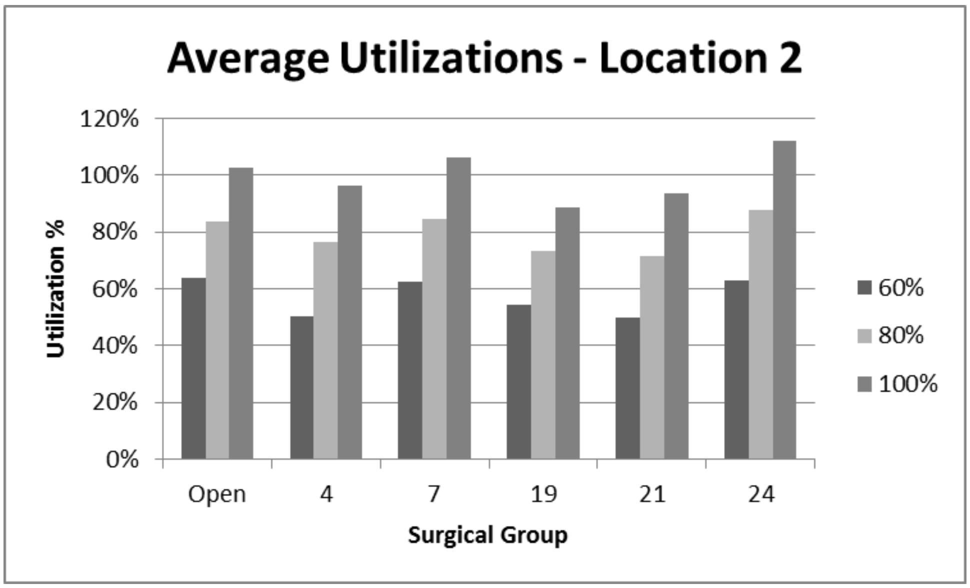

Figure 3.5: Average utilizations for expected utilizations using a Medium-Low averseness to OT

One thing to consider when adding in a buffer is how many cases are being pushed back or lost compared to the previous method. An average of four cases were pushed back or lost for each 20% buffer added to the original simulation. When deciding to add an extra case or use a buffer, managers should weigh the profits of each surgery against any potential overtime resource costs.

5. CONCLUSIONS AND FUTURE WORK

We built a more advanced simulation based on previous work from Chapter II involving analysis of block release policies. In addition, we used work from Chapter I to build the schedule run in the simulation model. Data from a local hospital was used to test three Averseness to OT scenarios and two block release policies.

Our simulation results showed little differences between averseness to OT levels but showed utilization increases for certain groups when switching between block release policies. When switching from a 3-day block release to a no release policy, groups with Medium classifications in location 1 saw the highest utilization increases. The other largest or smallest increases did not fit into similar classifications.

Additionally, we showed how using a buffer in the schedule can help reduce potential overtime in ORs. When using a buffer for scheduling, OR managers must consider costs between resource costs and surgical revenue.

As in Chapter I, we focus on assigning surgeries to rooms but do not require a specific order for the cases. Equipment limitations and block release times were also not considered for the open posting room cases, which can be complicated when considering the number of services that can use this time. While we showed high utilization levels, we must note that these are distributions and estimates of case arrivals, durations, and accuracy levels. One can also assume these results will be better than that what can be seen in practice.

CHAPTER V
CONCLUSIONS

This thesis contributes to the area of operations research and simulation in healthcare by providing detailed findings, flexible tools, and suggested policies about managing the OR suite. This research focuses on the strategic planning levels of setting up schedules and policies before the day of surgery occurs.

Chapter II presents work built on a previous mathematical model from Hosseini and Taaffe [6] to assign hours to surgical groups and introduces an additional model for allocating assigned time to specific ORs. By using this type of mathematical programming approach, we allow for a more generalized approach that focuses on the level of risk averseness in terms of overtime tolerance, rather than specifying and quantifying these costs for ORs. Additionally, we provide an open posting encouragement level that creates more shared, open blocks and uses fewer ORs. This tool can be used in environments with constrained resources or it can be used to develop and promote the idea of open posting time, which is currently uncommon but gaining acceptance in hospital settings. We support our models by providing a trace-driven simulation using data from a local hospital. We use 12 months of historical data to create the schedule and then "simulate" the next three months of historical data to determine the goodness of fit of the proposed schedule. Our results showed that using a lower averseness to overtime resulted in fewer assigned surgical hours, higher utilization levels, and more overtime, while using a higher averseness to overtime resulted in more assigned surgical hours, lower utilization levels, and less overtime. OR managers can select an averseness level that matches the hospital's environment and capacity for overtime. Averseness levels can also be assigned by surgical group, giving high averseness levels to services that are variable or expect growth and lower averseness levels to services that are more predictable. Our results also showed that using high open posting encouragement levels resulted in fewer scheduled ORs and more shared, open posting time, while using little to no open posting encouragement resulted in few changes to the schedule and more scheduled ORs. Again, OR managers can select an encouragement level that matches the hospital's political environment and resource constraints. If the hospital is introducing the concept of open posting time, a lower encouragement level may be more acceptable. This chapter provides methods that will allow managers flexibility in creating and managing the OR schedule while still obtaining desirable utilization levels.

Chapter III presents a detailed analysis of the effects of block release policies on room utilization levels and number of unscheduled cases. Surgery arrival rates, surgical case durations, and expected capacity are all varied and tested in order to provide insights about common block release policies. Two specific policies were tested and compared for a single blocked and open posting room. Our results showed strong impacts on utilization levels and number of unscheduled cases when viewing each room separately. However, since each room preferred a different block release policy, when these results are viewed as a whole, the impacts are very negligible. Overall, our results showed that forced block release benefits rooms that have lower utilization levels (around 60%). However, if the block is operating at such a low utilization, one must consider if their room hours should be reallocated. This chapter provides further insights from the work of Dexter and Macario [31] and can be used with hospital or surgical group-specific distributions to test different block release policies before they are put into effect in the real world.

Chapter IV presents the results of a case study that combines the work of Chapters II and III. We use data from the local hospital in Chapter II to create schedules with varying averseness to overtime levels, surgery arrival distributions, surgical duration estimates, and prediction accuracy levels. We then build on the simulation created in Chapter III to add ORs and specify their hours based on the schedules created. Our results showed little differences between release policies in schedules that varied in averseness to overtime levels. Differences were seen in the groups with medium classifications when switching from a 3-day release policy to a no release policy. We noted that our results showed very high utilization levels for all release policies that were most likely due to inaccurate surgical duration predictions, which were also seen in Chapter II. In order to view more realistic utilization levels, we provided a buffer that accounted for differences in expected versus actual surgical durations. This buffer showed more realistic utilization levels but increased the number of unscheduled surgeries. For this reason, OR managers must consider the tradeoffs between resource costs and surgical revenue. The findings of this chapter validate the use of the proposed mathematical models from Chapter II and allow for specific insights about block release policies based on a particular hospital's historical data.

REFERENCES

[1] Litvak E, Long MC. Cost and Quality under Managed Care: Irreconcilable Differences? American Journal of Managed Care 2000:6:305–312.

[2] McIntosh C, Dexter F, Epstein RH. Impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: tutorial using data from an Australian hospital. Anesthesia & Analgesia 2006; 103:1499–1516.

[3] Cardoen B, Demeulemeester E, Van der Hoeven J. On the use of planning models in the operating theatre: results of a survey in Flanders. International Journal of Health Planning and Management 2010; 25:400-414.

[4] Strum DP, Vargas LG, May JH, Bashein G. Surgical suite utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1997; 90: 1176–1185.

[5] Strum DP, Vargas LG, May JH, Bashein G. Surgical suite utilization and capacity planning: a minimal cost analysis model. Journal of Medical Systems 1997; 21: 309–322.

[6] Hosseini N, Taaffe KM. Allocating operating room block time using historical caseload variability. Healthcare Management Science 2014.

[7] Blake JT, Donald J. Mount Sinai Hospital uses integer programming to allocate operating room time. Interfaces 2002; 32:63–73.

[8] Dexter F, Blake J, Penning D, Lubarsky D. Calculating a potential increase in hospital margin for elective surgery by changing operating room time allocations or increasing nursing staffing to permit completion of more cases: a case study. Anesthesia & Analgesia 2002; 94:138–42.

[9] Kuo PC, Schroeder RA, Mahaffey S, Bollinger RR. Optimization of operating room allocation using linear programming techniques. Journal of the American College of Surgeons 2003; 197: 889–895.

[10] Mannino C, Eivind JN, Nordlander TE. A Pattern Based, Robust Approach to Cyclic Master Surgery Scheduling. Journal of Scheduling 2012; 15:553–563.

[11] Dexter F, Macario A. Changing Allocations of Operating Room Time From a System Based on Historical Utilization to One Where the Aim Is to Schedule as Many Surgical Cases as Possible Anesth Analg 2002; 94:1272–9.

[12] Kumar A, Shim S. Using Computer Simulation for Surgical Care Process Reengineering in Hospitals. INFOR 2005: 303–319.

[13] Ballard SM, Kuhl ME. The use of simulation to determine maximum capacity in the surgical suite operating room. In: Proceedings of the 38th Winter Simulation Conference 2006: 433–38.

[14] Dexter F, Macario A, Lubarsky DA. The impact on revenue of increasing patient volume at surgical suites with relatively high operating room utilization. Anesthesia & Analgesia 2001; 92:1215–1221.

[15] Testi A, Tanfani E, Torre G. A three-phase approach for operating theatre schedules. Health Care Management Science 2007; 10:163–172.

[16] Shamayleh A, Fowler J, Zhang M. Operating Room Capacity Planning Decisions. In: Proceedings of World Academy of Science, Engineering and Technology 2012:665-669.

[17] Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR. Optimal allocation of surgery blocks to operating rooms under uncertainty. Operations Research 2010; 58:802–816.

[18] Dexter F, Traub RD. How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesthesia & Analgesia 2002; 94:933-942.

[19] Denton BJ, Viapiano J, Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Management Science 2007; 10:13–24.

[20] Van Houdenhoven M, van Oostrum JM, Wullink G, Hans E, Hurink JL, Bakker J, Kazemier G. Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. Journal of Critical Care 2008; 23:222-226.

[21] Agnetis A, Coppi A, Corsini M, Dellino G, Meloni C, Pranzo M. A decomposition approach for the combined master surgical schedule and surgical case assignment problems. Health Care Management Science 2013; 17:49–59.

[22] Van Houdenhoven M, van Oostrum JM, Hans EW, Wullink G, Kazemier G. Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. Anesthesia & Analgesia 2007; 105:707-714.

[23] Lehtonen JM, Torkki P, Peltokorpi A, Moilanen T. Increasing operating room productivity by duration categories and a newsvendor model. International Journal of Health Care Quality Assurance 2013; 26: 80-92.

[24] Stepaniak PS, Mannaerts GHH, de Quelerij M, de Vries G. The effect of the operating Room coordinator's risk appreciation on operating room efficiency Anesthesia & Analgesia 2009; 108:1249-1256.

[25] Van Essen JT, Hurink JL, Hartholt W, van den Akker BJ. Decision support system for the operating room rescheduling problem. Health Care Management Science 2012; 15:355-372.

[26] Epstein RH, Dexter F. Statistical power analysis to estimate how many months of data are required to identify operating room staffing solutions to reduce labor costs and increase productivity. Anesthesia & Analg, 2002; 94:640-643.

[27] Green L. Queuing Analysis in Healthcare. In: Patient Flow: Reducing Delay in Healthcare Delivery. Springer-US, 2006:281–307.

[28] Dexter F, Epstein RD, Traub RD, Yiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. Anesthesiology 2004; 101:1444-1453.

[29] Wachtel RE, Dexter F. Review of behavioral operations experimental studies of newsvendor problems for operating room management. Anesthesia & Analgesia 2010; 110:1698-1710.

[30] Dexter F, Shi P, Epstein RH. Descriptive Study of Case Scheduling and Cancellations Within 1 Week of the Day of Surgery." Anesthesia & Analgesia 2012; 115:1188–1195.

[31] May JH, Spangler WE, Strum DP, Vargas LG. The Surgical Scheduling Problem: Current Research and Future Opportunities. Production and Operations Management 2011; 20:392–405.

[32] Dexter F, Macario A. When to Release Allocated Operating Room Time to Increase Operating Room Efficiency. Anesthesia & Analgesia 2004; 98:758-762.

[33] Patterson P. A few simple rules for managing block time in the operating room. In OR Manager 2004; 1–12.

[34] Dexter F, Traub RD, Macario A. How to Release Allocated Operating Room Time to Increase Efficiency: Predicting Which Surgical Service Will Have the Most Underutilized Operating Room Time. Anesthesia & Analgesia 2003; 96:507–512.

[35] Jun JB, Jacobson SH, Swisher JR. Application of Discrete-Event Simulation in Health Care Clinics: A Survey. The Journal of the Operational Research Society 1999; 50:109–123.

[36] Hamrock E., Kerrie P., Parks J, Scheulen J, Levin S. Discrete Event Simulation for Healthcare Organizations: A Tool for Decision Making. Journal of Healthcare Management 2013; 58:110-125.

[37] Dexter F, Macario A, Traub RD, Hopwood M, Lubarsky DA. An Operating Room Scheduling Strategy to Maximize the Use of Operating Room Block Time: Computer Simulation of Patient Scheduling and Survey of Patients' Preferences for Surgical Waiting Time." Anesthesia & Analgesia 1999; 89:7–20.

[38] Persson MJ, Persson JA. Analysing management policies for operating room planning using simulation. Health Care Management Science 2010;13:182–191.

[39] Macario AM. Are Your Hospital Operating Rooms "Efficient?": A Scoring System with Eight Performance Indicators. Anesthesiology 2006; 105:237–240.

[40] Davies R, Davies H. Modelling patient flows and resource provision in health systems. Omega 1994; 22:123-131.

[41] Steins K, Persson F, Holmer M. Increasing Utilization in a Hospital Operating Department Using Simulation Modeling. Simulation 2010; 86:463-480.

[42] Larsson A. The accuracy of surgery time estimations. Production Planning & Control: The Management of Operations 2013; 24:891-902.

[43] Zhou J, Dexter F, Macario A, Lubarsky DA. Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. Journal of Clinical Anesthesia 1999; 11:601-605.

[44] Dexter F, Traub RD, Qian F. Comparison of Statistical Methods to Predict the Time to Complete a Series of Surgical Cases. Journal of Clinical Monitoring and Computing 1999; 15:45-51.

[45] Dexter F, Dexter EU, Masursky D, Nussmeier NA. Systematic Review of General Thoracic Surgery Articles to Identify Predictors of Operating Room Case Durations. Anesthesia & Analgesia 2008; 106:1232-1234.

[46] Combes C, Meskens N, Rivat C, Vandamme JP. Using a KDD process to forecast the duration of surgery. International Journal of Production Economics 2008; 112:279-293.

[47] Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. Anesthesiology 2000; 92:1454-166.