Scheduling Elective Surgeries in Multiple Operating Rooms

Cansin Cagan Acarer

Submitted in partial fulfilment

of the requirements for the degree of

Master of Science in Management

(Operations and Information Systems Management)

Goodman School of Business, Brock University

St. Catharines, Ontario

Abstract

This thesis focuses on the problem of designing appointment schedules in a surgery center with multiple operating rooms. The conditions under which overlapping surgeries in the surgeons' schedule (i.e. parallel surgery processing) at the lowest cost are investigated with respect to three components of the total cost: waiting time, idle time, and overtime. A simulation optimization method is developed to find the near-optimal appointment schedules for elective surgical procedures in the presence of uncertain surgery durations.

The analysis is performed in three steps. First, three near-optimal operating room schedules are found for different cost configurations based on the secondary data of surgery durations obtained from the Canadian Institute for Health Information. Second, these near-optimal appointment schedules are used to test a parallel scheduling policy where each surgeon has overlapping surgeries scheduled in two operating rooms for the entire session (480 minutes) and only attends the critical portions of surgeries in the two operating rooms. Lastly, another parallel scheduling policy is tested where each surgeon has overlapping surgeries scheduled for half of the session duration (240 minutes) and only has surgeries scheduled in one operating room for the remaining time. These two policies are tested using simulation with scenarios for parallelizable portions of surgeries varying from 0.1 to 0.9 at 0.1 increments and three cost configurations.

In the simulated scenarios, the total cost is calculated as the weighted sum of patient waiting time, surgeon idle time, surgeon overtime, operating room idle time, and operating room overtime. Out of the nine scenarios for each policy and each cost configuration, the parallelizable portion of surgeries that result in the lowest total cost is identified.

The results from both policies indicate that implementing parallel scheduling policies for surgery types with higher parallelizable portions results in surgeons remaining idle for longer periods during the session. This idle time cost is justified by a decrease in other cost components for surgeries with parallelizable portions 50% or less; however, the total cost is higher for surgeries with parallelizable portions over 50%. In addition, it has been observed that overlapping surgeries with lower parallelizable portions is more expensive than overlapping those over with 50%. Therefore, it is concluded that the surgery types that allow parallel surgery scheduling policies to be implemented at the lowest cost have 50% of their duration parallelizable.

Keywords: parallel surgery scheduling, operating room scheduling, simulation optimization.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Dr. Reena Yoogalingam, for her mentorship, patience, and diligence. I am grateful for her continued support and guidance on my thesis and every other matter I asked help for since my first day in the program. She has been a true inspiration to me.

I would like to thank my thesis committee members, Dr. Anteneh Ayanso, Dr. Martin Kusy, and Dr. Esaignani Selvarajah. Their helpful comments greatly improved the quality of my thesis.

Finally, I would like to thank my parents. I could not be in this program without their financial and emotional support.

Table of Contents

List of Tables

List of Figures

1 Introduction

The health care industry is one of the major drivers of the economy in many industrialized countries. Health care spending represents, on average, 9% of GDP in the OECD countries (OECD 2017). This figure is expected to reach 14% by 2060 (OECD 2013) because in many countries health care costs are growing faster than the economies (OECD 2004). In an uncertain and changing industry, health care institutions are under increasing pressure to reduce costs and provide timely access to high quality services. Whether it is due to budget caps or for the purpose of staying competitive, health care institutions around the world are faced with the challenge of improving the quality of care to meet rising patient expectations while limiting their costs and managing their resources efficiently.

Providers of health care services include primary care clinics, specialty care clinics, outpatient procedure centers (also known as ambulatory surgery centers), and hospitals. In most OECD countries, hospitals account for the largest portion of current health expenditure (OECD 2017). They provide a variety of services including preventative care, emergency care, diagnostic services (e.g. examination, imaging, testing) as well as therapeutic services such as physical therapy and surgery. Providing these services requires hospitals to undertake large capital expenditures for facilities and equipment. In addition, salaried employees can be considered as a fixed cost in the short term (Roberts 1999). In order to remain efficient, hospitals have to operate these resources at a high level of utilization and distribute these costs over a large number of patients.

Operating Rooms (ORs) and related services are the largest cost center of a hospital (Macario, Vitez, Dunn, & McDonald 1995). The major components of the cost associated with ORs are human resources (such as the salaries of the surgeons, anesthesiologists, and other staff) and physical resources (such as ORs and specialized equipment). High variation in procedure durations and uncertainty in demand, makes it difficult to manage these resources efficiently.

One key determinant of OR efficiency is the surgery appointment schedule. The main decisions in an Appointment Scheduling Problem (ASP) are the allotted times for each appointment and the sequence in which the patients are served. A well-designed appointment schedule aims to reduce waiting time of patients while keeping resource utilization high and the overtime low. This is typically modeled as a weighted sum of patient waiting time, resource idle time, and overtime. Depending on the particular circumstances of the health care provider, these cost components have different levels of importance. In general, the cost coefficients for overtime and idle time of the resources are set significantly higher than the coefficient for patient waiting time to reflect the higher cost of OR resources.

In this study, the focus is on the problem of determining the appointment times for a given number of elective procedures in surgery centers with multiple ORs. The objective is to minimize the weighted sum of patient waiting time, resource idle time, and overtime costs. In addition, the impact of parallel surgery processing, which has received limited attention in the literature, is be investigated.

The ASP is a particularly difficult problem to solve because of the inherent complexity and variability that exists in the system. In the simplest case, it is assumed that patients have the same service time characteristics and they arrive punctually at their appointment times. Even this case involves a level of variability due to the uncertainty in service durations. The problem becomes more complex with the introduction of multiple patient types with different service time characteristics, multiple service stages with different resource requirements, and multiple doctors. OR scheduling is further complicated by the need to simultaneously schedule specialized providers and resources (such as surgeons with a particular specialty, specialized nursing staff, and equipment) according to the surgery type. Surgery appointments are longer and more variable. On a given day, many different types of procedures are performed in a surgery center.

The job allowance (or the allotted time) for a surgery is the difference between the start time of that surgery and the next surgery. The job allowance of each surgery is generally determined based on the observed mean duration for that procedure, but the realized surgery durations can fluctuate significantly. These fluctuations typically result either in idle periods during which the capacity is wasted, or in the OR running overtime at an additional cost. Patients may also not be ready before their scheduled appointment time even if resources become available earlier than expected. This may be because the patient has not been prepared to start the surgery, or in an outpatient setting, the patient is not present until the appointment time. Figure 1 illustrates a planned schedule of surgeries performed by the same surgeon in one OR and the impact of varying surgery durations on the realized schedule. The realized schedule shows that Surgery 1 finished earlier than expected. This resulted in idle time since it was not possible to start Surgery 2 before its scheduled start time. Surgery 2 is an example of a procedure that takes longer than expected. Even though the procedure time of Surgery 3 did not vary from expected duration, Surgery 2 caused Surgery 3 to be delayed, resulting in waiting time for Patient 3 and overtime for the OR.

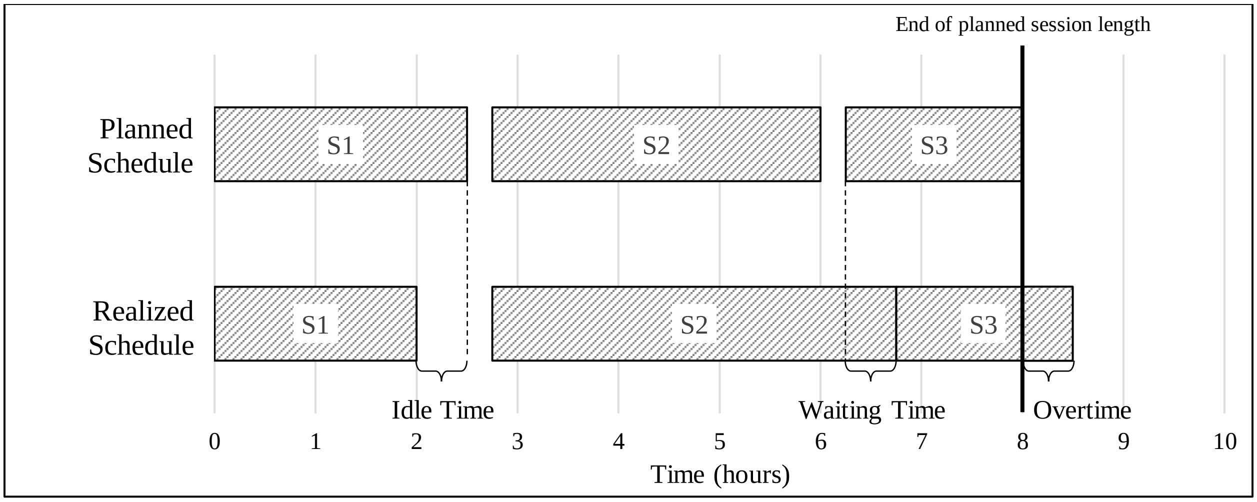

Figure 1. Planned and realized schedules of a single OR

Most studies in the literature focus on the OR scheduling problem for surgery centers with multiple ORs. A common method for creating OR schedules is block booking (Magerlein & Martin 1978). In this system, blocks of OR time are reserved for surgical specialties or individual surgeons. This simplifies the planning process because the problem comes down to two decisions: allocating the time blocks to the ORs and scheduling the surgeries within a block similar to single server ASP. However, block booking systems may lead to low OR utilization if the unused time in the blocks is not released for use by other surgical specialties. In open booking (also known as non-blocked) systems, an OR scheduler receives the OR time requirements from surgeons and creates the daily schedule for each OR (Erdogan & Denton 2011). Open booking systems are not favorable for specialties that cannot predict demand far in advance since the available OR time is booked on a first-come, first served basis in this system. Hybrid systems (Gupta & Denton 2008), also known as modified block systems (Patterson 1996), combine these two methods either by releasing the unused block time after a deadline, or by block booking only a portion of the available OR time and leaving the rest of the OR time available for open booking.

In block booking systems, surgeries in an OR time block are usually performed by one surgeon. As seen on Figure 1, variation in surgery duration only affects subsequent surgeries performed in the same OR. In open booking systems, surgeons are usually assigned cases in more than one OR on the same day. In this case, a delay in one surgery may have a compounded effect. Consider a case with 2 ORs, 3 surgeons, and 7 surgeries (S1, …, S7). Surgeon 1 performs S1, S2, and S7; Surgeon 2 performs S4 and S5; Surgeon 3 performs S3 and S6. Suppose the surgeries are scheduled as shown in Figure 2.

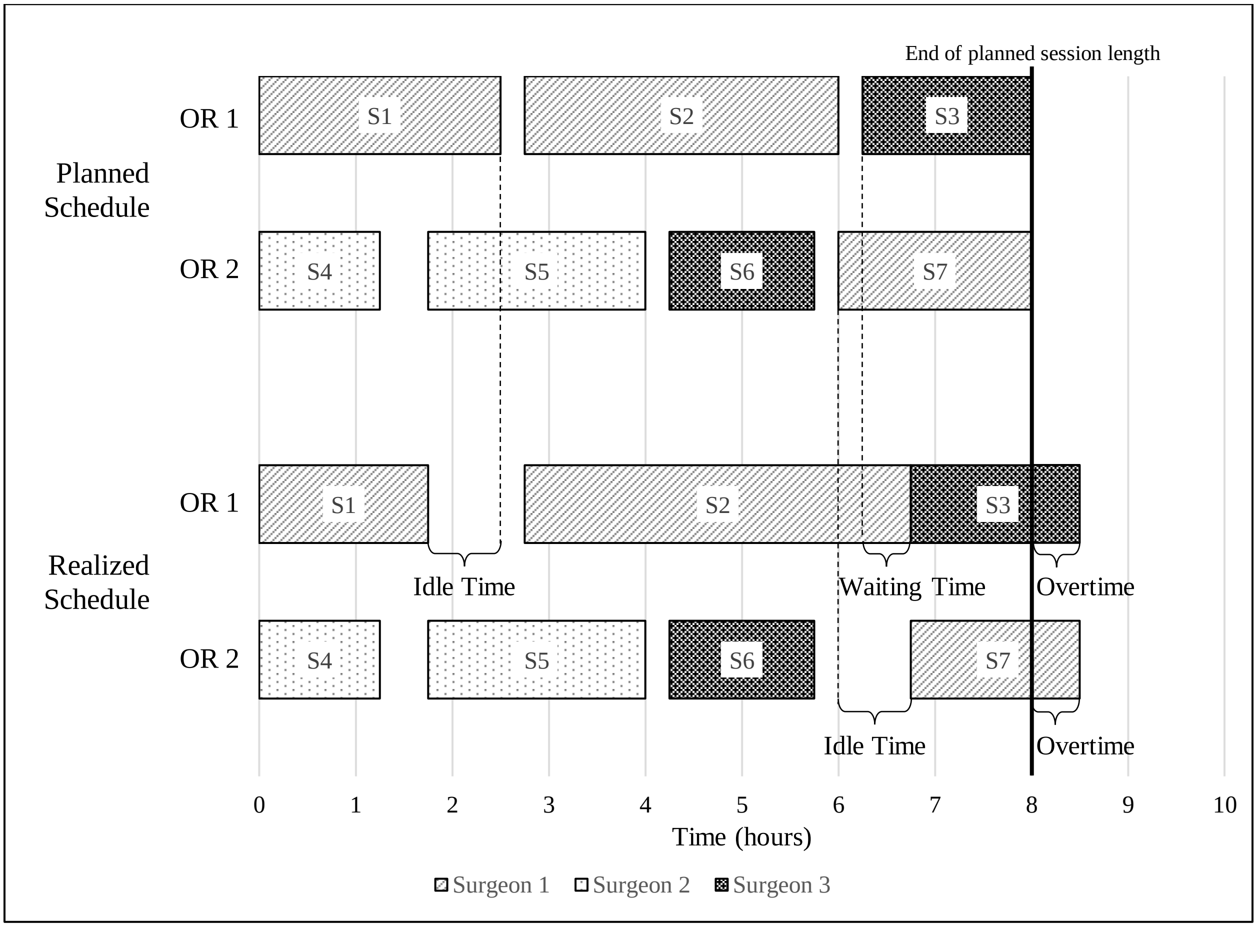

Figure 2. Planned and realized schedules of a surgery center with 2 ORs and 3 surgeons

Similar to the previous example, S1 finishes earlier than expected, S2 takes longer than its allotted time and the realized durations of other surgeries are the same as their respective scheduled durations. The delay of S2 not only leads to the delay of subsequent surgeries in OR 1, but also delays S7 in OR 2 because Surgeon 1 is still busy at the scheduled start time of S7. This results in both ORs running overtime. Since ORs are the most expensive resource in a hospital, any idle time and overtime result in a significant increase in cost. A carefully designed appointment schedule can offset these costs by minimizing the negative effects of this variability in surgery duration.

Designing OR schedules solely for the purpose of minimizing OR related costs may result in poor utilization of another expensive resource: surgeons' available time. More recently, a number of studies have explored the possibility of parallel surgery scheduling. This thesis builds on the approach that was proposed by Batun, Denton, Huschka, and Schaefer (2011). As illustrated in Figure 3, they modeled surgery process as three separate activities: preincision, incision, and postincision. They pointed out that surgeons are only required for the incision phase of the surgery and can be considered idle during other noncritical activities. In their model, they allowed consecutive surgeries in a surgeon's list to overlap for the noncritical portion of the surgeries as shown in Figure 3. They reported that this strategy leads to significant cost reduction.

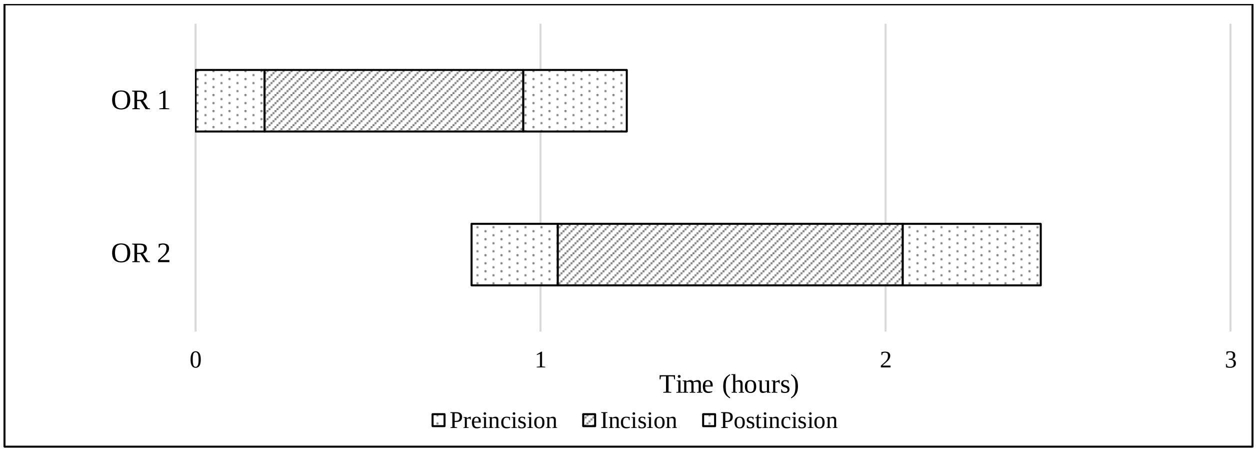

Figure 3. Parallel schedule in two ORs

Parallel surgery processing is widely used at academic medical centers and community hospitals to utilize lead surgeons' time more efficiently, therefore allowing more timely access to care (Ravi et al. 2018). It also provides learning experience under supervision for surgical trainees before they assume the responsibility of independent practice (Guan, Karsy, Brock, Couldwell, & Schmidt 2017). Although there is public concern about this practice (Boodman 2017), studies have shown that this practice has no significant effect on the outcome or post-operative complication rates for elective surgery patients (Liu et al. 2017; Zhang et al. 2016). Despite its significant benefits having been reported by a few studies, parallel surgery scheduling received very limited attention in the literature. This study aims to contribute to the literature by exploring the benefits of this approach further and determining the conditions under which it can be implemented in a viable way.

The complexity of this problem and the significant variability that exists necessitates the use of a method that can simultaneously address these aspects. In this thesis, a simulation optimization approach is used as the solution method. Simulation optimization is a stochastic optimization method. It has been used to solve problems that contain large complex search domains with many sources of uncertainty (Klassen & Yoogalingam 2008). The simulation component accounts for the stochastic parameters in the model. It allows uncertainties such as surgery durations to be represented in a wide range of probability distributions. Using probability distributions that best represent the real-life data results in a more realistic model and more applicable results. In their survey study, Guerriero and Guido (2011) conclude that the modeling flexibility makes simulation the most reliable and efficient tool to handle the complexity and variability of surgery planning and scheduling problems. Jun, Jacobson, and Swisher (1999) also point out that simulation is well-suited to tackle resource allocation problems in health care environments. In this thesis, simulation is combined with an optimization algorithm based on scatter search. Scatter search is a population-based metaheuristic optimization technique that is capable of handling the complex nature of the problem. In each iteration, the algorithm simultaneously searches different areas of the search domain and evaluates the candidate solutions using the simulation model. In addition, it utilizes tabu search and a neural network accelerator to improve the search efficiency. This study is the first to use this approach to study parallel surgery scheduling.

In their literature survey, Cayirli and Veral (2003) pointed out the need for studies with generalized applicability, as opposed to studies that analyze a specific clinic. In this study, data collected by the Canadian Institute for Health Information (CIHI) is obtained from their Discharge Abstract Database (DAD) and used to provide numerical examples for the simulation optimization models. The data comprises anonymized records of patient visits. It includes surgery duration records of a variety of different institutions and different surgery types. This ensures the generalized applicability of the results of this study and contributes to the body of empirical evidence on surgery scheduling.

2 Literature Review

There is a vast amount of research on the management of ORs in the operations research and management science literature. The OR scheduling problem investigated in this thesis is an application of the widely studied Appointment Scheduling Problem (ASP). In this section, the literature on OR scheduling studies is reviewed. In particular, those that involve scheduling elective surgical procedures, which are the focus of this thesis, are discussed. Since the OR scheduling problem is a variant of the ASP, the literature on the ASP problem is first briefly reviewed. The interested reader is referred to the reviews of Cayirli and Veral (2003) and Gupta and Denton (2008) for extensive reviews of the studies focusing on ASP in outpatient settings and other environments.

2.1 The Appointment Scheduling Problem

The objective of the ASP is to find an Appointment System (AS) that optimizes a particular performance measure (Cayirli & Veral 2003). Typically, the performance measure is the total cost which is calculated as the weighted sum of patient waiting time, resource idle time, and overtime. An AS specifies the intervals between appointment times (job allowances) and the number of patients to be scheduled for each appointment time (block size). If there are distinct patient characteristics in terms of service time, then determining the sequence of the appointments (i.e. the order in which different types are scheduled) becomes relevant. For environments where the frequency of no-shows and add-on cases (e.g. walk-ins, urgent and emergency cases) is significant, the AS needs to be adjusted to account for these factors. The overall goal of the research in this field is to develop generalized ASs that can be implemented in various settings by accounting for the relevant factors.

Before the ASP received the attention of researchers, the widespread appointment scheduling policy was the single block system. In this system, the patients were only given an appointment date and they were served on a First-Come, First-Served (FCFS) basis on the day of their appointments. This policy minimizes the resource idle time, but it also results in long patient waiting times. The individual block system in which patients were assigned unique appointment times was first analyzed by Bailey (1952). Assuming patient punctuality and the same distribution for service durations, he compared different appointment systems using simulation. He found that scheduling two patients at the beginning of the session and the rest of the patients with equal job allowances results in lower patient waiting times and an acceptable level of utilization. This is known as Bailey's Rule.

Most studies in the literature, aim to minimize the total cost by finding a good trade-off between the patient waiting time and resource idle time. Fetter and Thompson (1965) developed a simulation model of an outpatient clinic with unpunctual patients, no-shows, and walk-ins. They conclude that increasing the utilization from 60% to 90% is only worthwhile if the cost coefficient of resource idle time is ten times greater than the coefficient of patient waiting time. In their simulation study, Ho and Lau (1992) tested nine scheduling rules with 27 combinations of different no-show probabilities, service time variations, and number of patients per session. They found that Bailey's Rule performs the best when there is no information available about the cost of resource idling and patient waiting.

Later studies have shown that when service durations are independent and identically distributed (i.i.d.), the optimal job allowances exhibit a dome shape (Denton & Gupta 2003; Jiang, Tang, & Yan 2019; Robinson & Chen 2003; Wang 1993), rather than the equally sized job allowances proposed by Bailey (1952). The Dome Rule sets the appointment times with increasing intervals until the middle of the session, then sets the rest of the appointments with decreasing intervals. Klassen and Yoogalingam (2009) showed that setting job allowances equal for the appointments in the middle of the day provides more robust solutions across different performance measures and cost coefficients at a variety of settings. These studies assume that either the sequence of appointments is predetermined or the patient characteristics are homogeneous (i.e. only one type of patient is scheduled); therefore, the sequence of the appointments does not affect the results. This is a realistic assumption for environments such as primary care settings where the variation of case durations tend to be low (Gupta & Denton 2008). However, case durations are more variable for surgery appointments.

Cayirli, Veral, and Rosen (2006) used simulation of clinic sessions to experiment with six environmental factors (the number of patients scheduled per clinic session, the probability of no-shows, the probability of walk-ins, the coefficient of variation for service times for two patient types, and the mean patient punctuality), six sequencing rules, and seven appointment rules. Using the simulation results, they investigated the main and interaction effects of these factors on the cost of waiting time and combined cost of idle time and overtime. They found that the sequencing rule has a greater effect on these performance measures compared to the appointment rule.

Patient types are relevant for scheduling purposes because they usually translate into different probability distributions for service durations. When the problem involves patient groups with different service time characteristics, the sequence of the patients is relevant for scheduling purposes. This problem is first addressed by Weiss (1990), who solved the sequencing and scheduling problem for two patients. He showed that the solution highly depends on the cost coefficients.

In primary care environments, the appointment scheduling problem can be considered as a single server problem because the physicians have separate queues and they serve patients in their own office. However, the OR scheduling problem is more complex because ORs are usually shared by multiple surgeons.

2.2 The OR Scheduling Problem

In the appointment scheduling studies described above, the objective is to minimize the total cost of patients' waiting time, physician's idle time and overtime. Since physician's idle time comes at a higher cost relative to patient waiting time, keeping the physician highly utilized is the more important concern for scheduling and sequencing decisions. Similarly, the objective of the OR scheduling problem is to minimize the total cost of patient waiting time, resource idle time, and overtime. However, scheduling surgery appointments is more complex because compared to the service durations of primary care consultations, the surgery durations are longer and they are more variable since complications are far more likely. Moreover, OR scheduling involves simultaneously scheduling other expensive resources such as ORs, equipment, and supporting staff to be available at the start time of each case. Therefore, OR scheduling studies typically include the idle time and overtime costs for a combination of these resources.

In block booking environments where a surgeon is assigned a block of time within a particular OR, the idle times and overtimes of a surgeon and their assigned OR happen at the same time; meaning when the surgeon is idle the OR is idle and when the surgeon is working overtime, the OR is running overtime. The problem of scheduling and sequencing the surgeries within a time block is similar to the ASP of a primary care physician who works from the same office for the entire day. Both problems can be modeled using one idle cost and one overtime cost for the physician and the other resources. In an open booking system, surgeons do not have their surgeries scheduled back to back in the same OR; therefore, their schedule and the costs associated with it have to be modeled separately from the ORs they work in. The focus of most studies is on surgery centers with multiple ORs. The OR scheduling problem in a multi-OR setting involves the additional decision of allocating the surgeries to ORs. In general, most studies approach this problem in one of two ways: block booking and open booking systems.

In block booking systems, the first step is the decision of allocating the time blocks to the ORs, the second step is to schedule surgeries within the block. Since the second step is very similar to the ASP in primary care environments, early analytical studies that consider the multi-OR setting focused on the first step. A number of studies refer to the first step problem as the OR planning problem. The resulting schedule that shows the allocation of OR time blocks to the surgical specialties over a certain time horizon is referred to as the master surgical schedule (MSS).

Blake and Donald (2002) and Santibáñez, Begen, and Atkins (2007) considered the problem of designing MSS for a weekly time horizon. Blake and Donald (2002) formulated a mixed-integer programming (MIP) model to determine the weekly schedule of blocks that minimizes the undersupply of OR time for each surgical specialty. Santibáñez et al. (2007) modeled this problem for a system of hospitals with the objective of maximizing the total throughput of patients and minimizing the maximum post-surgical resources used by one surgical specialty. They formulated an MIP that determines the OR each surgical specialty is assigned and the number of procedures to be performed on each day. In contrast to Blake and Donald (2002), they constrained the problem so that each OR can be assigned to one surgical specialty on a given day. Testi, Tanfani, and Torre (2007) proposed a three-step approach. For the first two steps, they proposed one Integer Programming (IP) model to determine the number of cases from each surgical specialty to be scheduled over the horizon, and another IP model to determine the weekly schedule. For the third step, they tested different sequencing rules considering the urgency and priority of patients. They found shortest processing time (SPT) to result in the lowest number of delayed and postponed operations.

In block booking systems, the surgery appointment time is determined in two steps from the patient's point of view. First, the patient selects a time window from the available time blocks allocated to the particular surgical specialty. A time block is considered full when the total estimated durations of the scheduled surgeries reach the session length. Once the schedule is determined, the patient is notified of the exact appointment time. These steps are referred to as advanced scheduling and allocation scheduling (Cardoen, Demeulemeester, & Beliën 2009; Magerlein & Martin 1978).

In general, studies that focus on open booking systems first assign the surgeries to the ORs, then each OR is scheduled as a single server problem, similar to the problem of scheduling surgeries within a block. The study of Guinet and Chaabane (2003) is an early example of an analytical study that approaches the multi-OR problem in two steps. They formulated an MIP for assigning surgeries to ORs under resource constraints. They proposed two solution methods: The first is to relax certain constraints to make their problem equivalent to the assignment problem so that it can be solved with the Hungarian method optimally. The second is a primal-dual heuristic which is shown to perform well in their experiments. They proposed the second step problem of sequencing the individual ORs to be solved by an extension of the Gilmore and Gomory (1964) algorithm. Jebali, Hadj Alouane, and Ladet (2006) formulated one MIP that minimizes the total cost of overtime, OR idle time, and patient waiting time for assigning surgeries to ORs and another MIP for the sequencing of surgeries in each OR. They also tested a strategy that allows OR assignments to be changed in the second step but they found this strategy does not always provide a better solution despite being computationally more challenging. Fei, Meskens, and Chu (2010) proposed a similar two-step approach. They first assign each surgery a date by solving a set partitioning IP model using a column-generation-based heuristic. Then they formulated the problem of determining the daily schedule of ORs and recovery rooms as a two-stage hybrid flow-shop problem and proposed a solution with a hybrid genetic algorithm. These studies did not incorporate any uncertainty in their models.

Lamiri, Xie, Dolgui, and Grimaud (2008) studied the problem of determining the set of elective surgeries to be performed in each period over a time horizon in the presence of uncertain demand for emergency cases. They used Monte Carlo simulation to account for the uncertainty of emergency demand and solve an MIP with the sampled values. They showed that this Monte Carlo Optimization method yields lower costs compared to solving the same MIP deterministically using expected emergency demand. They also proved that as the computation budget increases (i.e. as the number of samples is increased), their solution converges to the optimal. In a later study, Lamiri, Grimaud, and Xie (2009) compared several heuristic methods for solving the same problem. They found that even the heuristic that performed best (Tabu Search) is outperformed by their Monte Carlo Optimization method. In another study, Lamiri, Xie, and Zhang (2008) extended their earlier work and considered the assignment of elective cases to ORs. They formulated a stochastic linear programing model and proposed a column generation approach to find a near-optimal solution for realistic sized problem instances. Although uncertainty of emergency demand is considered, these studies assume surgery durations to be deterministic.

Stochastic surgery durations were first used in simulation studies. An early example is the study of Barnoon and Wolfe (1968). They generated probability distributions from actual data and sampled from these distributions in their scenarios. They tested the performance of a multi-OR setting under different resource levels. A more recent example is the study of Wullink et al. (2007). Assuming lognormal surgery durations, they compared the policy of running a dedicated OR for emergency cases against reserving capacity in all ORs. Their simulation results indicated the latter to perform better.

A number of analytical studies also modeled stochastic surgery durations. Gerchak, Gupta, and Henig (1996) proposed a stochastic dynamic programming model to determine the number of elective cases to be booked under the uncertain demand of emergency cases. Hans, Wullink, van Houdenhoven, and Kazemier (2008) compared different heuristics for assigning elective surgeries to ORs and minimizing the slack for uncertain surgery durations. Denton, Viapiano, and Vogl (2007) relaxed the assumption of i.i.d. surgery durations from their earlier study (Denton & Gupta 2003) and proposed a stochastic optimization model. They did not provide an exact solution method, but they tested 3 heuristics (sequencing surgeries in a block in the order of increasing mean of durations, variance of durations, and coefficient of variation durations) on the deterministic equivalent of their model. Using data from a health care center, they provided numerical experiments and showed that these heuristics perform better than the schedules used in the subject hospital.

Dexter, Macario, and Traub (1999) used simulation to test 10 different scheduling algorithms for add-on cases. They found that assigning cases to ORs in descending order of scheduled duration while allowing scheduled surgeries to exceed the session end time provides the highest OR utilization. Bam, Denton, Oyen, and Cowen (2017) developed two different two-stage heuristics for block scheduling systems. Their “fast two-phase heuristic” first assigns each surgeon’s block using LPT and sequences the surgeries with a heuristic method. Using simulation, they compared this method to an MIP model they proposed and they showed that the heuristic model performs well with the additional advantage of being easy to implement.

Only a few studies in the literature address OR scheduling problem using simulation-optimization approaches. Lin, Sir, and Pasupathy (2013) proposed a multi-objective simulation-optimization framework for determining the number of nurses, anesthetists, and pre-operative beds that minimize the average patient waiting time and system completion time. In the proposed framework, they first used genetic algorithm to generate values for the level of these resources. Then, they estimated the patient waiting time and system completion time using simulation. Lastly, they used Data Envelopment Analysis (DEA) to evaluate these performance measures. They demonstrated the efficiency of their framework with the numerical results of a case study.

Denton, Rahman, Nelson, and Bailey (2006) proposed a simulated annealing algorithm to determine patient arrival schedules that minimize the total cost of overtime and patient waiting time. Using a numerical example, they showed that their method for schedule optimization provides 50% improvement in the total cost compared to the existing schedule.

Denton, Miller, Balasubramanian, and Huschka (2010) proposed models with deterministic and stochastic surgery durations for the problem of assigning surgeries to ORs. First, they formulated the problem with deterministic surgery durations as a variant of the extensible bin packing problem. Then, they extended it into a two-stage stochastic linear program with the binary decisions of opening each OR in the first stage, and the overtime decisions for each OR after the uncertain surgery durations are observed in the recourse function. Extending the model proposed by Denton et al. (2010), Batun et al. (2011) added the decisions of allocating surgeries to ORs, setting the precedence relationships, and determining the start times in the first stage of the problem. They pointed out the inefficiencies of block booking; they considered OR time as a shared resource and allowed surgeons to be assigned surgeries in more than one OR on a given day. Their results show that this practice leads to a cost reduction between 22% and 59% on average depending on the cost of surgeon idle time. In addition, they modeled the surgery as a three-step process (pre-incision, incision, post-incision) where the primary staff surgeon is only busy during one step (incision). This allowed them to schedule another surgery for the primary staff surgeon in a different OR after his or her turnover time, but before the first surgery ends. They referred to this practice as parallel surgery processing.

2.3 Parallel Processing in the Literature

In a manufacturing context, the term parallel processing refers to a variant of multiple-machine scheduling problem where a set of jobs is processed by a set of machines and the objective is to minimize the time it takes to finish processing all jobs or minimize the cost of missing due dates. In this problem, parallel machine scheduling refers to a setting where a job can be processed in any one of the free machines and leave the system (Cheng & Sin 1990). The problem in this setting is to allocate machines to jobs and to set the sequence of processing. Unlike the OR scheduling problem, the only resource being considered in this problem is the machine.

In the parallel processing approach described in this thesis, ORs are the equivalent of machines in the parallel machine scheduling problem but in this study, surgeons are also considered as another set of resources. This is similar to the dual resource constrained (DRC) problems in manufacturing which involves labor as an additional constraint. This problem involves developing rules for the assignment of labor to the machines (Ramasesh 1990), in addition to the aforementioned decisions in the parallel machine scheduling problem. However, OR scheduling differs from this problem because each job (surgery) in a surgery center comes with a predetermined surgeon assigned to it which makes the problem less flexible.

In the context of OR management, parallel surgery processing has received limited attention in the literature. Different studies use this term to refer to different practices depending on which processes they proposed to run in parallel. Hanss et al. (2005) and Torkki, Marjamaa, Torkki, Kallio, and Kirvelä (2005) studied cases where the anesthesia induction is performed before the preceding procedure ends. Even though this practice requires additional resources (another room and another team for the anesthesia induction), both studies conclude that the financial benefit of increasing the utilization of the most expensive resource (OR) outweighs the cost of these additional resources. Berg et al. (2010) used simulation to test varying levels of resources at a colonoscopy suite. They found that as the ratio of the number of procedure rooms to the number of endoscopists increases (up to the ratio of 2:1), the patient throughput increases because this enables endoscopists to serve patients in parallel.

In this thesis and in the study of Batun et al. (2011), the term parallel processing is used to describe a different setting where surgeries use two resources that operate independently. The first is the surgeon who is not required for all the activities during the time patient spends in the OR with other members of the surgical team. The second is the OR and the other members of the surgical team whose schedule is assumed to be aligned with the time patient spends in the OR. The term parallel processing in this thesis refers to the practice of a surgeon leaving one OR after completing their activities and starting a new surgery in another OR before the first surgery is completed by the first surgical team. Parallel surgery processing is also referred to as concurrent or overlapping surgeries in the medical literature (Beasley et al. 2015; Zhang et al. 2016).

3 Problem Formulation and Methodology

The base model is introduced in section 3.1.1. This is the general form of the ASP from the literature. In section 3.1.2, this formulation is extended to model multiple ORs. This model is extended further in section 3.2 to include three stages of the surgical procedure and parallel processing of the non-critical stages. The methodology used to solve this problem is described in section 3.2. A brief description of the data set used for the numerical experiments is provided in section 3.3.

3.1 Model Formulation

3.1.1 Base Model

In its simplest form, the appointment scheduling problem is one of finding the optimal vector of appointment times $\boldsymbol{X}^* = [x_1, \dots, x_N]^{\mathrm{T}}$ to minimize the weighted sum of the expected costs of total waiting time, total idle time, and overtime in the presence of uncertain appointment durations. In this problem, it is assumed that the patients have the same service time characteristics.

The following problem instance-related parameters are assumed to be known in advance:

N: Number of surgeries (i.e. patients) to be scheduled.

d: Planned session length in minutes.

P: Random variable representing the appointment duration.

c_W: Cost coefficient for patient waiting time.

c_I: Cost coefficient for idle time.

c_O: Cost coefficient for overtime.

The decision variable:

x_i: Appointment time (scheduled start time) for patient i for $i \in \{1, \dots, N\}$.

The appointment durations in the simulation model are represented with the following expression.

$p_i(\omega)$: Realized duration of appointment i sampled from P in scenario ω

$i \in \{1, \dots, N\}$.

The following notation is used for the performance measures.

$W_i(\omega)$: Waiting time of patient i in scenario ω.

$I_i(\omega)$: Idle time immediately before patient i in scenario ω.

$O(\omega)$: Overtime in scenario ω.

In this thesis, ω is used as the scenario index. For example, $O(\omega)$ represents the overtime observed as the result of the random appointment durations realized in scenario (i.e. simulation replication) ω. $E[O]$ is the expected value of overtime estimated by taking the average of $O(\omega)$ values obtained from a sufficient number of replications. It is assumed that the first appointment starts at the beginning of the session ($x_1 = 0$); therefore, $W_1(\omega) = 0$ and $I_1(\omega) = 0$. The performance measures for the other surgeries are defined as shown below.

$$W_i(\omega) = max[x_{i-1} + W_{i-1}(\omega) + p_{i-1}(\omega) - x_i, 0] \qquad for\ i \in \{1, \dots, N\} \tag{1}$$

$$I_i(\omega) = max\{x_i - [x_{i-1} + W_{i-1}(\omega) + p_{i-1}(\omega)], 0\} \qquad for\ i \in \{1, \dots, N\} \tag{2}$$

$$O(\omega) = max[x_N + W_N(\omega) + p_N(\omega) - d, 0] \tag{3}$$

The model is formulated with the objective function defined as the weighted sum of the performance measures:

$$\min_{x_i} c_W E\left[\sum_{i=2}^{N} W_i(\omega)\right] + c_I E\left[\sum_{i=2}^{N} I_i(\omega)\right] + c_O E[O(\omega)] \tag{4}$$

$$s.t.\ 0 \le x_i \le d\ \ \forall i \tag{5}$$

$$x_1 \le x_2 \le \cdots \le x_N \tag{6}$$

$$x_i\ integer \tag{7}$$

3.1.2 Extended Model for Multiple Surgeons and Multiple ORs

The mathematical model presented in this section extends the base model to account for the decisions of allocating surgeries to operating rooms in a multiple operating room setting. The following problem instance-related parameters are assumed to be known in advance.

M: Number of surgeons to be scheduled.

N_m: Number of surgeries (i.e. patients) to be scheduled for surgeon m. $m \in \{1, \dots, M\}$

K: Number of ORs available.

P: Random variable representing the appointment duration.

d: Planned session length in minutes.

c_W: Cost coefficient for patient waiting time.

c_I: Cost coefficient for OR idle time.

c_O: Cost coefficient for OR overtime.

The surgery index i is assumed to be assigned such that surgeries of Surgeon 1 are given indices from 1 to N_1, surgeries of Surgeon 2 are given indices from $N_1 + 1$ to $N_1 + N_2$, and so on. The decision variables:

x_i: Appointment time of surgery i. $i \in \{1, \dots, \sum_m N_m\}$

y_i: OR number of the room where surgery i is allocated. $i \in \{1, \dots, \sum_m N_m\}$

The following expressions are derived from x_i and y_i in the simulation model. They are used in the calculation of the performance measures.

N_k^R: The number of surgeries allocated in OR k. $k \in \{1, \dots, K\}$

h_i^S: The rank of surgery i in the surgery list of the surgeon who will perform it. $i \in \{1, \dots, \sum_m N_m\}$

h_i^R: The rank of surgery i in the sequence of surgeries in the OR where it is allocated. $i \in \{1, \dots, \sum_m N_m\}$

u_i^S: The index of the surgery that immediately precedes surgery i in the list of the same surgeon. $i \in \{1, \dots, \sum_m N_m\}$

u_i^R: The index of the surgery that immediately precedes surgery i in the sequence of the same OR. $i \in \{1, \dots, \sum_m N_m\}$

The following notation is used for the performance measures.

$W_i(\omega)$: Waiting time of patient i in scenario ω.

$I_i^R(\omega)$: Idle time of the OR and the surgical team immediately before patient i in scenario ω.

$I_i^S(\omega)$: Idle time of the surgeon immediately before patient i in scenario ω.

$O_k^R(\omega)$: Overtime of the OR and the surgical team in scenario ω.

$O_m^S(\omega)$: Overtime of the surgeon m in scenario ω.

In this model, the schedule of a surgeon is not necessarily the same as the schedule of an OR throughout the session. For example, if an OR is used by one surgeon for the first half of the session and by another surgeon for the remaining time, the OR may be idle during the time between the completion of the last surgery of the first surgeon and the first surgery of the second surgeon. Since the idle time before a surgeon's first surgery is assumed to be 0, this time would not be considered as a surgeon idle time. To account for such cases, the idle time is calculated separately for surgeons and operating rooms.

The overtime is calculated separately for ORs and surgeons to account for feasible solutions where overtime of surgeons may be different from the overtime of ORs. Consider a problem instance with two surgeons and two operating rooms where both surgeons have their last surgery scheduled at the end of the session ($x_{N_1} = x_{N_1+N_2} = d$). If these surgeries are allocated to different ORs ($y_1 \neq y_2$), then the overtime of each surgeon would be equal to the overtime of the operating room in which they are performed. For this case, one set of overtime performance measures for either one of the resource types (ORs or surgeons) would be sufficient. However, if both surgeries are assigned to the same OR ($y_1 = y_2$), then only one of the ORs would be open beyond the session length while both surgeons work overtime. Such cases require overtime to be measured separately for ORs and surgeons.

The realized appointment durations in the scenario ω of the simulation model are represented with the following expression.

$p_i(\omega)$: Realized duration of surgery i sampled from P in scenario ω.

$i \in \{1, \dots, \sum_m N_m\}$

Once the surgery durations are sampled and the uncertainty is resolved for a scenario, start time of surgery i [$s_i(\omega)$] and the performance measures are calculated with the following formulations.

$$s_i(\omega) = \begin{cases} x_i & h_i^S = 1 \wedge h_i^R = 1 \\ \max\,[x_i, x_{u_i^R} + W_{u_i^R}(\omega) + p_{u_i^R}(\omega)] & h_i^S = 1 \wedge h_i^R \neq 1 \\ \max\,[x_i, x_{u_i^S} + W_{u_i^S}(\omega) + p_{u_i^S}(\omega)] & h_i^S \neq 1 \wedge h_i^R = 1 \\ \max\,[x_i, x_{u_i^R} + W_{u_i^R}(\omega) + p_{u_i^R}(\omega), x_{u_i^S} + W_{u_i^S}(\omega) + p_{u_i^S}(\omega)] & h_i^S \neq 1 \wedge h_i^R \neq 1 \end{cases} \quad (8)$$

$$i \in \{1, \dots, \sum_m N_m\}$$

$$W_i(\omega) = s_i(\omega) - x_i \tag{9}$$

$$I_i^R(\omega) = \begin{cases} \max\left\{s_i(\omega) - \left[s_{u_i^R}(\omega) + p_{u_i^R}(\omega)\right], 0\right\} & h_i^R \neq 1 \\ 0 & h_i^R = 1 \end{cases} \quad i \in \{1, \dots, \textstyle\sum_m N_m\} \tag{10}$$

$$I_i^S(\omega) = \begin{cases} \max\left\{s_i(\omega) - \left[s_{u_i^S}(\omega) + p_{u_i^S}(\omega)\right], 0\right\} & h_i^S \neq 1 \\ 0 & h_i^S = 1 \end{cases} \quad i \in \{1, \dots, \textstyle\sum_m N_m\} \tag{11}$$

$$O_k^R(\omega) = \max\left[s_{N_k}(\omega) + p_{N_k}(\omega) - d, 0\right] \qquad k \in \{1, \dots, K\} \tag{12}$$

$$O_m^S(\omega) = \max\left[s_{N_m}(\omega) + p_{N_m}(\omega) - d, 0\right] \qquad m \in \{1, \dots, M\} \tag{13}$$

This model is optimized for the weighted cost of waiting time and OR performances measures.

$$\min_{x_i, y_i} c_W E\left[\sum_{i=1}^{\sum_m N_m} W_i(\omega)\right] + c_I E\left[\sum_{i=1}^{\sum_m N_m} I_i^R(\omega)\right] + c_O E\left[\sum_{k=1}^{K} O_k^R(\omega)\right] \tag{14}$$

$$s.t.\ 0 \leq x_i \leq d \quad \forall i \tag{15}$$

$$x_i = 0 \quad \forall i \in \{1, 1 + N_1, 1 + \textstyle\sum_{m=1}^{2} N_m, 1 + \sum_{m=1}^{3} N_m, \dots, 1 + \sum_{m=1}^{M-1} N_m\} \tag{16}$$

$$x_{u_i^S} \leq x_i \quad \forall i \in \{1, \dots, \textstyle\sum_m N_m\} \backslash \{1 + N_1, 1 + \sum_{m=1}^{2} N_m, 1 + \sum_{m=1}^{3} N_m, \dots, 1 + \sum_{m=1}^{M-1} N_m\} \tag{17}$$

$$1 \leq y_i \leq K \quad \forall i \tag{18}$$

$$x_i, y_i\ integer \tag{19}$$

For patients who are the first in a surgeon's list, constraint (16) sets the appointment times to 0. For patients who are not the first in a surgeon's list, constraint (17) ensures that their appointment time is set no later than the next patient in that surgeon's list. Constraint (18) and (19) ensures that surgeries are only assigned to available ORs; constraint (19) also ensures integer appointment times.

3.1.3 Extended Model for Multiple Surgeons, Multiple ORs, and Parallel Surgery Processing

In this section, the formulation in section 3.1.2 is extended to model the pre-incision, incision, and post-incision activities separately. In addition to the problem instance-related parameters defined in section 3.1.2, the following expression is assumed to be known in advance in this model.

q: Parallelizable portion of surgery duration.

The decision variables (x_i and y_i) and the expressions derived from the decision variables (N_k^R, h_i^S, h_i^R, u_i^S, u_i^R, v_i^S, and v_i^R) which are defined in section 3.1.2 also apply to this model. The durations of surgery activities are defined as shown below.

$t_i(\omega)$: Realized total duration of appointment i sampled from P in scenario ω.

$pre_i(\omega)$: Realized pre-incision duration of appointment i derived from $t_i(\omega)$ and q in scenario ω.

$p_i(\omega)$: Realized incision duration of appointment i derived from $t_i(\omega)$ and q in scenario ω.

$post_i(\omega)$: Realized post-incision duration of appointment i derived from $t_i(\omega)$ and q in scenario ω.

The durations of surgery activities are calculated as shown below.

$$pre_i(\omega) = t_i(\omega) \times q/2 \tag{20}$$

$$p_i(\omega) = t_i(\omega) \times (1-q) \tag{21}$$

$$post_i(\omega) = t_i(\omega) \times q/2 \tag{22}$$

In the model described in section 3.1.2, it is assumed that patient waiting time begins if the resources are not ready at the appointment time. In contrast, this model allows pre-incision to begin even if the surgeon is not available, provided the OR and the surgical team is available. Therefore, the realized pre-incision start time differs from the start time formulation in section 3.1.2 and it only depends on surgery sequence in the respective OR. It is independent of h_i^S. The formulation for the start times of pre-incision, incision, and post-incision activities are given below.

$$s_i^{pre}(\omega) = \begin{cases} x_i & h_i^R = 1 \\ \max\,[x_i, s_{u_i^R}^{post}(\omega) + post_{u_i^R}(\omega)] & h_i^R \neq 1 \end{cases} \quad (23)$$

$$i \in \{1, \dots, \textstyle\sum_m N_m\}$$

$$s_i^{p}(\omega) = \begin{cases} s_i^{pre}(\omega) + pre_i(\omega) & h_i^S = 1 \\ \max\,[s_i^{pre}(\omega) + pre_i(\omega), s_{u_i^S}^{p}(\omega) + p_{u_i^S}(\omega)] & h_i^S \neq 1 \end{cases} \quad (24)$$

$$i \in \{1, \dots, \textstyle\sum_m N_m\}$$

Since there are not any additional resources required for the post incision activity, it is assumed to start immediately after incision is completed, without any waiting time between these activities. Start time of post incision is equal to the completion time of incision:

$$s_i^{post}(\omega) = s_i^{p}(\omega) + p_i(\omega) \qquad i \in \{1, \dots, \textstyle\sum_m N_m\} \quad (25)$$

This difference in start time calculation is due to the fact that when parallel surgery processing is allowed, a surgeon is only required for the incision and considered idle during pre-incision and post-incision activities. Consequently, the idle time and overtime of surgeons differ from those of the OR and the rest of the surgical team. Surgeons' idle time and overtime are calculated separately for surgeons to capture this difference. In this model, the same notation from section 3.1.2 is used for these measures $[I_i^R(\omega), I_i^S(\omega), O^R(\omega), O^S(\omega)]$ but waiting time is captured with two measures. When parallel surgery processing is allowed, patient waiting may occur at two different points in time. Waiting time before pre-incision begins unless the OR and the rest of the surgical team are available at the appointment time. If the surgeon is still busy when the pre-incision is completed, then patient waiting time before incision begins. The following notation is used to describe these waiting time measures.

$W_i^R(\omega)$: Waiting time of patient i before pre-incision in scenario ω.

$W_i^S(\omega)$: Waiting time of patient i before incision in scenario ω.

The performance measures are calculated using the following equations.

$$W_i^R(\omega) = s_i^{pre}(\omega) - x_i \tag{26}$$

$$W_i^S(\omega) = s_i^p(\omega) - \left[s_i^{pre}(\omega) + pre_i(\omega)\right] \tag{27}$$

$$I_i^R(\omega) = \begin{cases} \max\left\{s_i^{pre}(\omega) - \left[s_{u_i^R}^{post}(\omega) + post_{u_i^R}(\omega)\right], 0\right\} & h_i^R \neq 1 \\ 0 & h_i^R = 1 \end{cases} \quad i \in \{1, \dots, \textstyle\sum_m N_m\} \tag{28}$$

$$I_i^S(\omega) = \begin{cases} \max\left\{s_i^p(\omega) - \left[s_{u_i^S}^p(\omega) + p_{u_i^S}(\omega)\right], 0\right\} & h_i^S \neq 1 \\ 0 & h_i^S = 1 \end{cases} \quad i \in \{1, \dots, \textstyle\sum_m N_m\} \tag{29}$$

$$O_k^R(\omega) = \max\left[s_{N_k}^{post}(\omega) + post_{N_k}(\omega) - d, 0\right] \qquad k \in \{1, \dots, K\} \tag{30}$$

$$O_m^S(\omega) = \max\left[s_{N_m}^p(\omega) + p_{N_m}(\omega) - d, 0\right] \qquad m \in \{1, \dots, M\} \tag{31}$$

The objective function of this model is the weighted cost of waiting times and OR performances measures.

$$\min_{x_i, y_i} c_W E\left[\sum_{i=1}^{\sum_m N_m} W_i^R(\omega)\right] + c_W sE\left[\sum_{i=1}^{\sum_m N_m} W_i^S(\omega)\right] + c_I E\left[\sum_{i=1}^{\sum_m N_m} I_i^R(\omega)\right] + c_O E\left[\sum_{k=1}^{K} O_k^R(\omega)\right] \quad (32)$$

$$s.t.\ 0 \le x_i \le d \quad \forall i \quad (33)$$

$$x_i = 0 \ \ \forall i \in \{1, 1 + N_1, 1 + \textstyle\sum_{m=1}^{2} N_m\,, 1 + \sum_{m=1}^{3} N_m\,, \dots, 1 + \sum_{m=1}^{M-1} N_m\} \quad (34)$$

$$x_{u_i^S} \le x_i \ \ \forall i \in \{1, \dots, \textstyle\sum_m N_m\} \backslash \{1 + N_1, 1 + \sum_{m=1}^{2} N_m\,, 1 + \sum_{m=1}^{3} N_m\,, \dots, 1 + \sum_{m=1}^{M-1} N_m\} \quad (35)$$

$$1 \le y_i \le K \quad \forall i \quad (36)$$

$$x_i, y_i\ integer \quad (37)$$

The roles of the constraints in this model are identical to the ones described in section 3.1.2.

3.2 Simulation Optimization Approach

Solving this problem with analytical techniques would require the closed form of the objective function or further simplifications of the model such as assuming the random variables to be exponentially distributed. Since the performance measures in the objective function are functions of stochastic appointment durations, $p_i(\omega)$, the closed form of the objective function can only be written with probability density function of P. This would make the problem intractable to solve with analytical techniques. Instead, numerical optimization methods are preferable because they are able to perform optimization with only the decision variable values and a mechanism to estimate the objective function's value (Gosavi 2015).

In this thesis, the simulation optimization algorithm embedded in OptQuest (OptTek Systems, Inc. 2005) is used in combination with a discrete-event simulation models developed in Arena 10.0 (Rockwell Software, Inc. 2005) and Monte Carlo simulation models developed in @Risk (Palisade Corporation 2002). The metaheuristic in OptQuest is capable of searching for $\boldsymbol{X}^*$ without knowing the structure of the objective function (Laguna & Wubbena 2005). The value of the objective function is estimated with a simulation model that takes the decision variables and the probability distribution of appointment durations as inputs and outputs a statistic for the objective function value. As pointed out by Law & Kelton (1991), simulations that use random variables as inputs produce random variables as outputs. Therefore, a sufficient number of simulation runs (replications) are performed to estimate the objective function value for a given set of decision variable values. This method allows a wide range of probability distributions to be used for the random variables.

Simulation optimization is used to address problems in a wide range of fields from manufacturing (Irizarry, Wilson, & Trevino 2001) and inventory management (Tsai & Chen 2017) to epidemiology (Ferris, Deng, Fryback, & Kuruchittham 2005). For the surgery scheduling problem, Saremi, Jula, ElMekkawy, and Wang (2013) proposed three simulation-based optimization methods. The method they reported to yield quality solutions with a short computational time involves generating solutions to a binary linear programming model using tabu search under the assumption of deterministic appointment durations. Then the resulting solution is evaluated with stochastic appointment durations in a discrete event simulation model developed in Arena 12. Several other studies used different methods combining simulation and optimization to solve the surgery scheduling problem (Lamiri et al. 2008; Lin et al. 2013; Zhang & Xie 2015). The particular approach used in this thesis has been established by Klassen and Yoogalingam (2009) as a robust solution method that is capable of finding good solutions for ASPs with a variety of different performance measures.

OptQuest was developed by Glover, Kelly, and Laguna (1996). It is a general-purpose optimizer and it treats the simulation model as a black box that outputs the objective function value for a given candidate solution vector ***X*** (i.e. reference point) (Kleijnen & Wan 2007; Laguna 1997a). It searches for the optimal solution with a metaheuristic based on scatter search (Glover 1977), tabu search (Glover 1986), and a neural network accelerator. In each iteration, OptQuest generates a population of candidate solutions and these are evaluated by calculating the corresponding objective function values in the simulation model (see Figure 4).

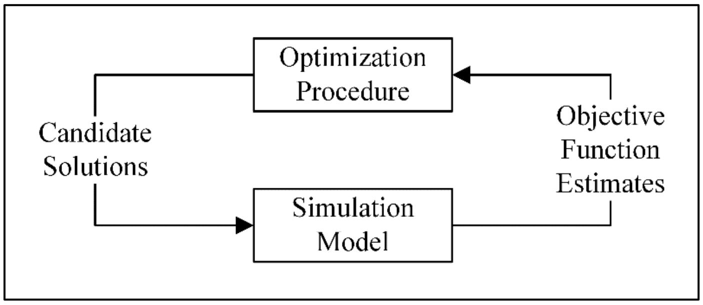

Figure 4. Coordination between simulation and optimization (adapted from Fu 2002)

The algorithm starts by generating a diverse population of initial candidate solutions. This population always includes a candidate solution in which the value of each decision variable is the midpoint between its predetermined lower bound and upper bound. Then each candidate solution is tested for feasibility by checking whether it satisfies the constraints. An infeasible reference point $\boldsymbol{X}$ is mapped to a point $\boldsymbol{X'}$ in the feasible region. $\boldsymbol{X'}$ is found by solving an MIP model that minimizes the absolute deviation between $\boldsymbol{X}$ and $\boldsymbol{X'}$.

Candidate solutions are evaluated by running the simulation model for a sufficient number of replications until a good estimate of the objective function value is obtained for each candidate solution. Since this is a computationally expensive process, OptQuest employs a neural network accelerator and an inferior solution test to limit the number of calls to the simulation model. The neural network is trained with an automatically determined number of reference points; then, it is used to predict the objective function value before the simulation is run or before the predetermined number of replications is reached (Laguna 1997b). An inferior solution test constructs a 95% confidence interval around the mean objective function value for the current candidate solution. If the best objective function value found using the candidate solutions tested up to that point does not fall into this interval, the current candidate solution is deemed inferior without running any other replications after the predetermined minimum number of replications is reached (Rockwell Automation 2012). Both good and inferior solutions are used to develop the next generation of solutions.

The next generation of candidate solutions are created using a scatter search heuristic. Based on their age (the number of iterations they have remained in the population) and the quality of their objective function value, two candidate solutions are selected as parent reference points to create four new reference points (offspring). The worst parent is replaced by the best offspring and the surviving parent is given a tabu-active status to prevent it from being selected as parent for a number of subsequent iterations. While increasing the quality of solutions in the population, this process may result in many reference points with similar characteristics. To ensure diversity in the population, the restarting procedure periodically injects newly created reference points to the population. The size of the population is determined by the system based on the time it takes to evaluate one candidate solution. This iterative process continues until convergence is reached (i.e., there is no statistically significant improvement in performance). A summary of the simulation optimization algorithm is given in Figure 5.

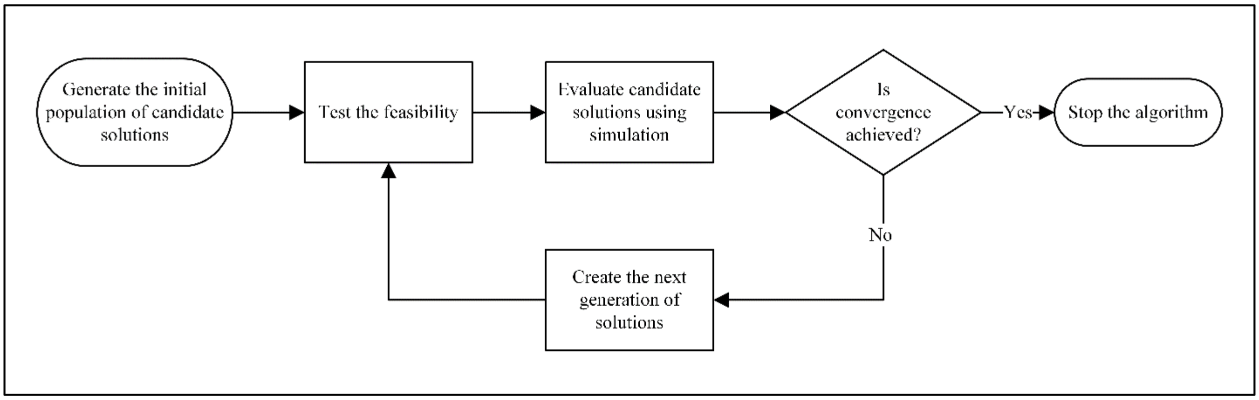
Generate the initial population of candidate solutions
Test the feasibility
Evaluate candidate solutions using simulation
Is convergence achieved?
Yes
Stop the algorithm
No
Create the next generation of solutions

Figure 5. Summary of the simulation optimization algorithm

3.3 Data

For the numerical experiments of this thesis, secondary data is obtained from the Discharge Abstract Database (DAD) of the Canadian Institute for Health Information (CIHI). CIHI is a non-profit organization that collects data from health care institutions across Canada. They maintain a number of databases about health services, health care personnel and health spending. In addition to the publicly available reports and analyses they publish; they also provide specific data from their databases for researchers and decision makers.

In their data quality report for the period in which our data was collected, CIHI (2017) reports that the percentage of invalid, missing or unknown values in their database is extremely low and the rate of duplicate records is 0.002%. The accuracy of the DAD data is also confirmed by Richards, Brown, and Homan (2002).

The dataset used in this thesis is comprised of day surgery records of a large health care institution for the period between April 2016 and March 2017. Data elements that contain identifying information such as institution and province names are received in deidentified form. The data elements include CCI code that describes the intervention type and anonymized codes for institution, the province where the institution is located, year the record is submitted, date and time when patient is admitted, date and time when intervention begins and ends. The full list of the obtained data elements and their descriptions are provided in Appendix A.

3.3.1 Data Cleaning

From the initial dataset, the records with missing or invalid values for Intervention Episode Start Time and Intervention Episode End Time are removed from the dataset. Since the data consists of outpatient surgery visits, the records with mismatching Discharge Date and Admission Date values are also assumed invalid and eliminated. Based on the empirical data in the literature it is found that surgery durations are never below 20 minutes for the surgery types in this dataset. Therefore, records with surgery durations below 20 minutes are also considered invalid and removed. Surgery durations are derived from the remaining 4177 records using the data elements named Intervention Episode Start Time and Intervention Episode End Time. Summary statistics for the surgery duration are given in Table 1 and the histogram of the surgery duration is given in Figure 6.

Table 1

Summary statistics for the surgery duration in minutes

Mean	37.45463251
Standard Error	0.295790796
Median	30
Mode	30
Standard Deviation	19.11687468
Sample Variance	365.4548977
Kurtosis	5.689335086
Skewness	2.117577728
Range	158
Minimum	20
Maximum	178
Count	4177

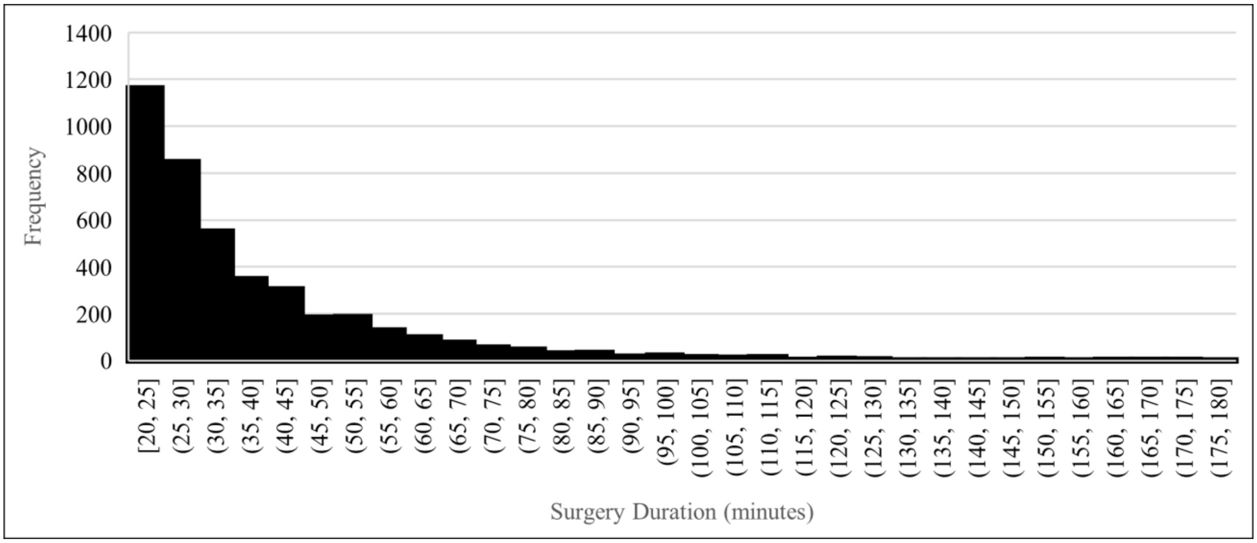

Figure 6. Histogram of surgery duration

3.3.2 Fitting a Probability Distribution to the Surgery Durations in the Dataset

Surgery durations are assumed to follow lognormal distribution in accordance with many studies in the literature (Jebali & Diabat 2017; May, Strum, & Vargas 2000; Zhang, Murali, Dessouky, & Belson 2009). Based on the assumptions described in section 3.3.1, the probability distribution to represent the random surgery durations in the simulation model is determined to be 20 + LOGN(18, 20). The mean of the values generated with this distribution does not have a statistically significant difference with the records in the clean dataset.

3.4 Experimental Design

First, the near-optimal schedules for an OR are obtained to be used as inputs for the experiments of different OR usage schemes. Based on the expected surgery durations in the dataset, number of patients to be scheduled in a 480 minutes OR session is determined as 12. In accordance with the literature, OR overtime is given 1.5 times more weight than that of OR idle time in all scenarios. Additionally, a coefficient that applies to both OR idle time and overtime is varied as 1, 20, and 50 to investigate the effects of patient waiting time having relatively less importance on the near-optimal schedule. The resulting values of the cost coefficients in the three objective functions are given in Table 2.

Table 2

Cost coefficient sets used in the objective functions

Cost Scenario #	c_W	c_I	c_O
1	1	1	1.5
2	1	20	30
3	1	50	75

Then, best schedules obtained under these three objectives are then used in the second set of experiments. These experiments are aimed to identify the conditions under which parallel surgery processing is favorable under two OR usage policies: (1) one surgeon using two ORs, (2) two surgeons each using one OR and sharing another. These two policies are tested in 54 scenarios by combining three different OR operating costs and values from 0.1 to 0.9 with 0.1 increments are tested for the parallelizable portion of surgeries (q) to identify the q values where the cost is lower. In line with the prior work in the literature (Batun et al. 2011), selecting these q values allows every procedure type that have any portion of its duration parallelizable to be covered exhaustively. The experimentation parameters are summarized in Figure 7.

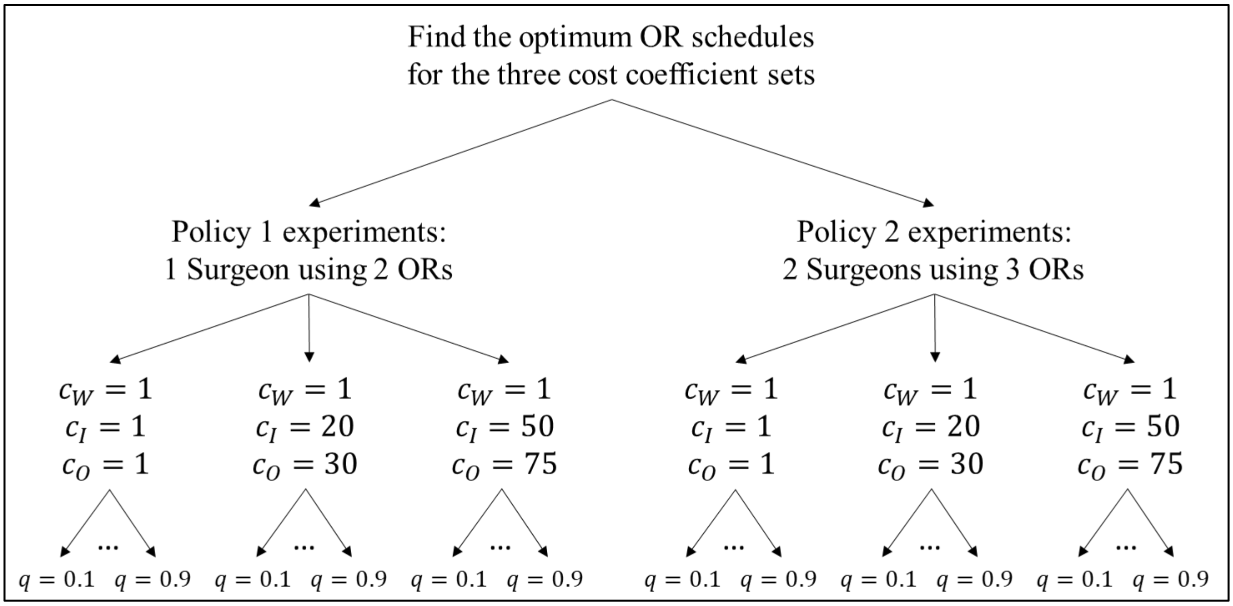
Find the optimum OR schedules for the three cost coefficient sets
Policy 1 experiments: 1 Surgeon using 2 ORs
Policy 2 experiments: 2 Surgeons using 3 ORs
$c_W = 1$ $c_I = 1$ $c_O = 1$
$c_W = 1$ $c_I = 20$ $c_O = 30$
$c_W = 1$ $c_I = 50$ $c_O = 75$
$c_W = 1$ $c_I = 1$ $c_O = 1$
$c_W = 1$ $c_I = 20$ $c_O = 30$
$c_W = 1$ $c_I = 50$ $c_O = 75$
…
$q = 0.1$
$q = 0.9$

Figure 7. Parameters tested at different levels of the experiments

4 Results and Analysis

4.1 Developing Near-Optimal Surgery Appointment Schedules

Three simulation optimization model is run with the cost coefficients given in Table 2. For each cost scenario, 100,000 candidate solutions are tested, and the objective function is estimated with 500 simulations for each solution. 100,000 candidate solutions are found to be more than sufficient since no improvements in the objective function greater than 0.001 are made after the 10,000th iteration. The resulting near-optimal job allowances shown on Figure 8.

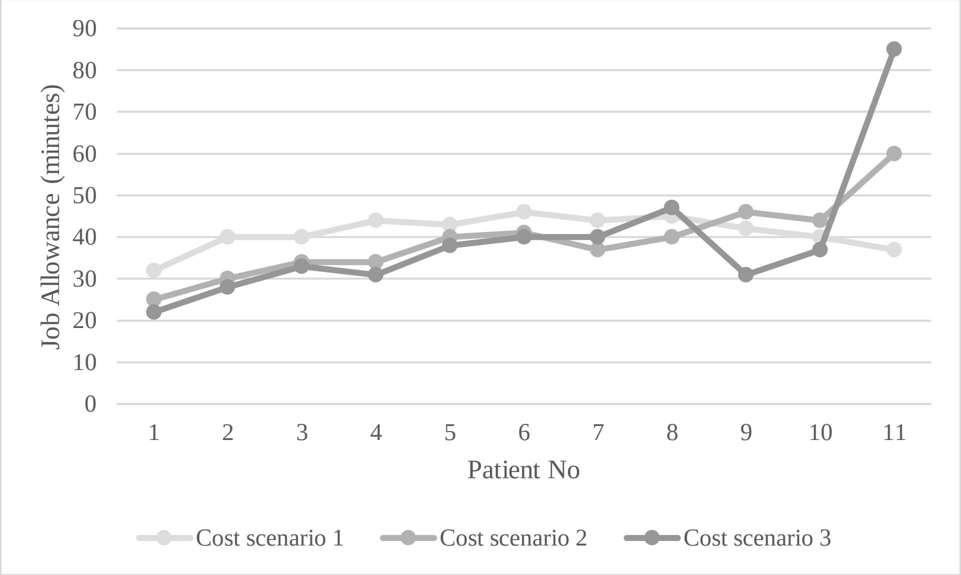

Figure 8. Near-optimal job allowances

These near-optimal job allowances are consistent with the previous studies mentioned in the literature review (Denton & Gupta, 2003; Jiang et al., 2019; Robinson & Chen, 2003; Wang, 1993) since they exhibit a dome shape. With higher cost of overtime, cost scenarios 2 and 3 exhibit higher job allowances for the last scheduled patient in the near-optimal solutions to avoid the cost of overtime. As expected, all near-optimal solutions set the appointment time of the first patient as 0. The near-optimal appointment times for all patients are given in Table 3.

Table 3

Near-optimal appointment times

Patient No (i)	**Appointment Times** (x_i) **in Cost Scenario 1** $c_W = 1$ $c_I = 1$ $c_O = 1$	**Appointment Times** (x_i) **in Cost Scenario 2** $c_W = 1$ $c_I = 20$ $c_O = 30$	**Appointment Times** (x_i) **in Cost Scenario 3** $c_W = 1$ $c_I = 50$ $c_O = 75$
1	0	0	0
2	32	25	22
3	72	55	50
4	112	89	83
5	156	123	114
6	199	163	152
7	245	204	192
8	289	241	232
9	334	281	279
10	376	327	310
11	416	371	347
12	453	431	432

Based on the total cost formula given in Expression (4), these near-optimal OR schedules yield the average total costs of 302.383, 1946.728, and 4319.332 for the three cost scenarios, respectively. The breakdown of these costs is shown on Figure 9 in terms of waiting time, idle time, and overtime. As expected, as the relative cost of idle time and overtime increases, the best schedule assigns appointment times closer to one another. This results in higher waiting time for the patients. As discussed in section 3.1, idle time and overtime are the same for surgeons and ORs when the surgeons remain in the OR for the entire duration of the surgery; therefore, they are not reported separately for each resource in this figure. In the following sections, these cost components are reported separately for surgeons and ORs.

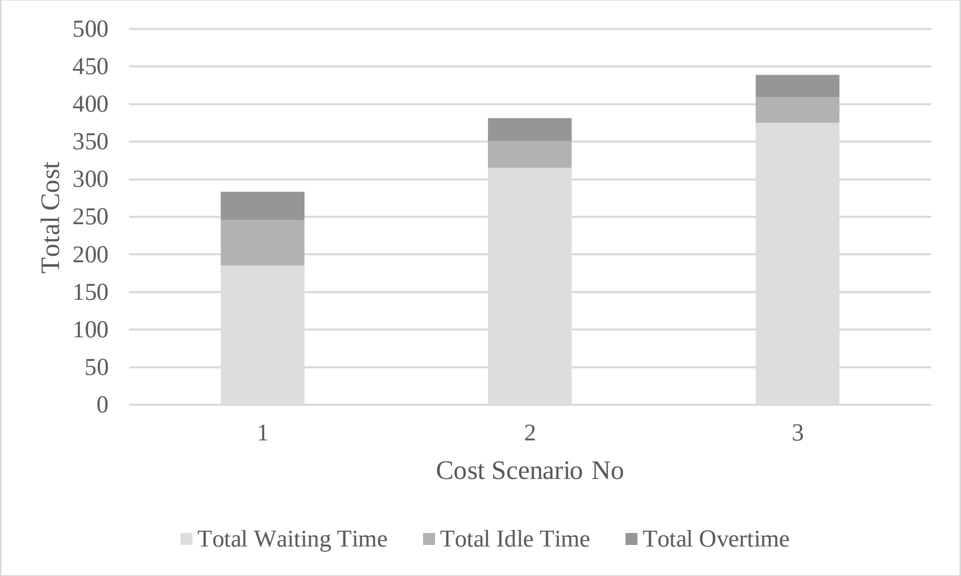

Figure 9. Total cost breakdown for the near-optimal OR schedule

4.2 Testing Parallel Surgery Scheduling Policies

In this section, two OR usage policies are tested to identify the conditions under which parallel surgery scheduling yields lower total cost. As summarized in Table 4, each policy is tested with three sets of cost coefficients in the weighted total cost calculation using the near-optimal appointment schedule found for those cost coefficients in section 4.1.

Table 4

Summary of simulation scenarios

Simulation Scenario* Parameters		Cost Coefficient Set			Near-Optimal Schedule Used
Policy No	Cost Configuration	c_W	c_I	c_O	
1	1	1	1	1.5	Cost Scenario 1
1	2	1	20	30	Cost Scenario 2
1	3	1	50	75	Cost Scenario 3
2	1	1	1	1.5	Cost Scenario 1
2	2	1	20	30	Cost Scenario 2
2	3	1	50	75	Cost Scenario 3

*Each combination of policy and cost configuration is tested with nine scenarios for $q = 0.1, 0.2, \dots, 0.9$

4.2.1 Policy 1: 1 Surgeon Using 2 Operating Rooms

The first scheduling policy investigated in this thesis is the situation where a surgeon who performs 12 surgeries in an OR for the entire session (480 minutes) is assigned a second OR with 12 additional surgeries. In this case, the surgeon only attends the critical portion (incision) of the surgeries in each OR. The first OR is scheduled based on the near-optimal schedule found in section 4.1. The schedule of the surgeries in the second OR is shifted by 19 minutes (half the expected duration of the surgery) forward in time from the near-optimal schedule so that the critical portions of the surgeries in the second OR are scheduled in between the ones in the first OR. A deterministic example of the scenarios tested for this case is illustrated on Figure 10. Using the notation from the mathematical model described in section 3.1, the resulting values of the other parameters for this example is given in Appendix B.

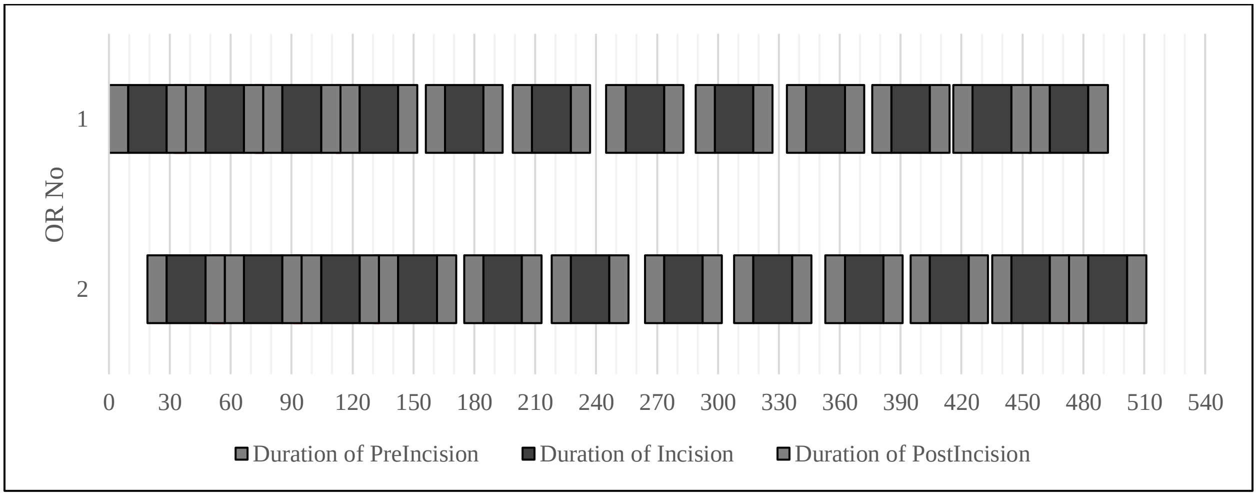

Figure 10. A deterministic example of Policy 1 scenarios

Under each of these cost configurations, nine scenarios are tested for the different values of the parallelizable portion of surgeries from 0.1 to 0.9. Each scenario is simulated for 5000 replications with Monte Carlo simulation. The resulting mean values and 95% confidence intervals for each of the cost components are reported in Appendix C. The expected value of the weighted total cost is calculated with the expression (38) using the respective cost coefficient sets for each of the simulation scenarios as described in Table 4. Hereafter, this output is referred to as OR Cost.

$$c_W E\left[\sum_{i=1}^{\sum_m N_m} W_i^R(\omega)\right] + c_W E\left[\sum_{i=1}^{\sum_m N_m} W_i^S(\omega)\right] + c_I E\left[\sum_{i=1}^{\sum_m N_m} I_i^R(\omega)\right] + c_O E\left[\sum_{k=1}^{K} O_k^R(\omega)\right] \tag{38}$$

The expected OR cost values found are shown by varying levels of the parallelizable portion of surgeries ($q = 0.1, 0.2, \dots, 0.9$) on Figure 11, Figure 12, and Figure 13 for cost configurations 1,2, and 3, respectively.

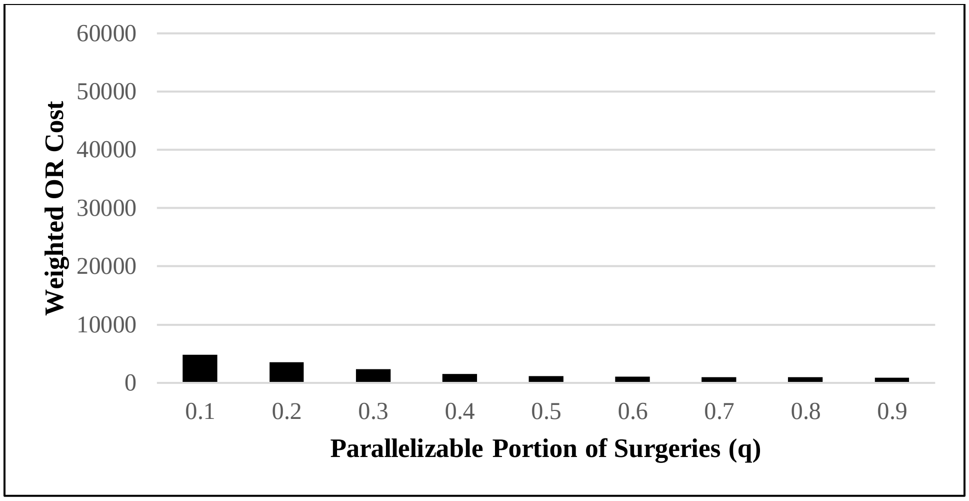

Figure 11. Total weighted OR cost for Cost Scenario 1

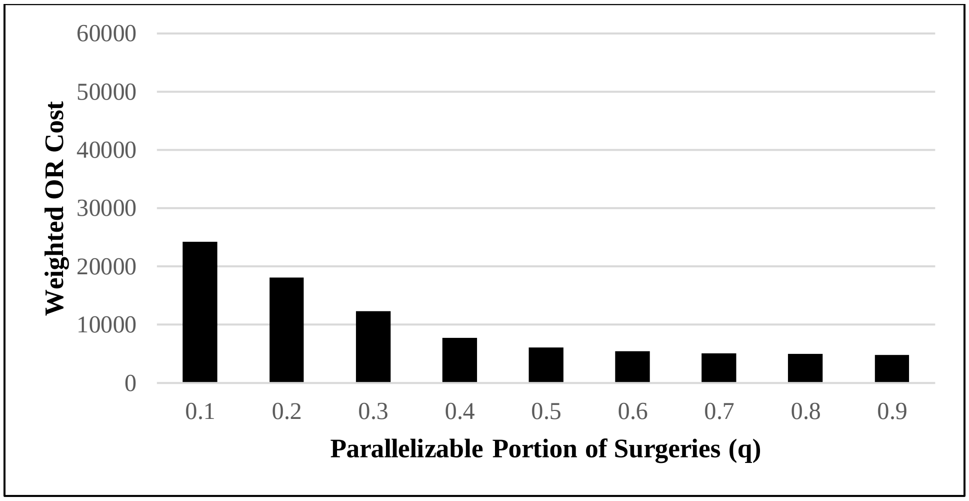

Figure 12. Total weighted OR cost for Cost Scenario 2

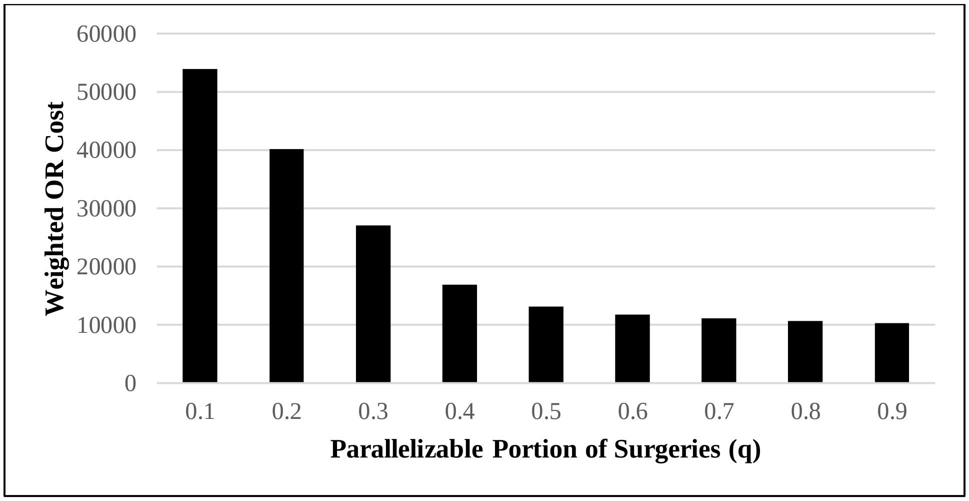

Figure 13. Total weighted OR cost for Cost Scenario 3

These results indicate that opening another OR and scheduling parallel surgeries for the surgeon in the second OR yields lower OR cost as the parallelizable portion of surgeries increases. This finding is in line with previous research in parallel surgery scheduling (Batun et al. 2011) and the numerical examples presented here shows that this result is true under different cost configurations.

For q values less than 0.5, the expected duration of the critical portion of surgeries in the second OR is greater than the expected time between the critical portions of two surgeries in the first OR. In other words, the critical portions of the surgeries are expected to overlap. Since the surgeon is required in the entire duration of the critical portion of a surgery, these overlaps disrupt the schedule in both ORs, and this problem is observed as higher costs for these q values consistently at different cost configurations. Therefore, it is less favorable to schedule surgeries with parallelizable portions less than 0.5 in parallel at two ORs.

In further analysis, surgeon idle time $I_i^S(\omega)$ and overtime $O_m^S(\omega)$ are taken into account for the total cost calculation. The weighted total cost is calculated using the Expression (39) and the respective cost coefficient sets for each of the simulation scenarios as described in Table 4. In contrast with the conclusions based solely on OR costs, the total cost does not continuously decrease as the parallelizable portion of surgeries increases if surgeon idle time and overtime is considered.

$$\begin{aligned} c_W E\left[\sum_{i=1}^{\sum_m N_m} W_i^R(\omega)\right] + c_{W^S} E\left[\sum_{i=1}^{\sum_m N_m} W_i^S(\omega)\right] + c_I E\left[\sum_{i=1}^{\sum_m N_m} I_i^R(\omega)\right] + c_O E\left[\sum_{k=1}^{K} O_k^R(\omega)\right] \\ + c_I E\left[\sum_{i=1}^{\sum_m N_m} I_i^S(\omega)\right] + c_O E\left[\sum_{k=1}^{K} O_k^S(\omega)\right] \end{aligned} \tag{39}$$

The total cost breakdown for varying levels of the parallelizable portion of surgeries ($q = 0.1, \ldots, 0.9$) are illustrated on Figure 14, Figure 15, and Figure 16 for cost configurations 1,2, and 3, respectively.

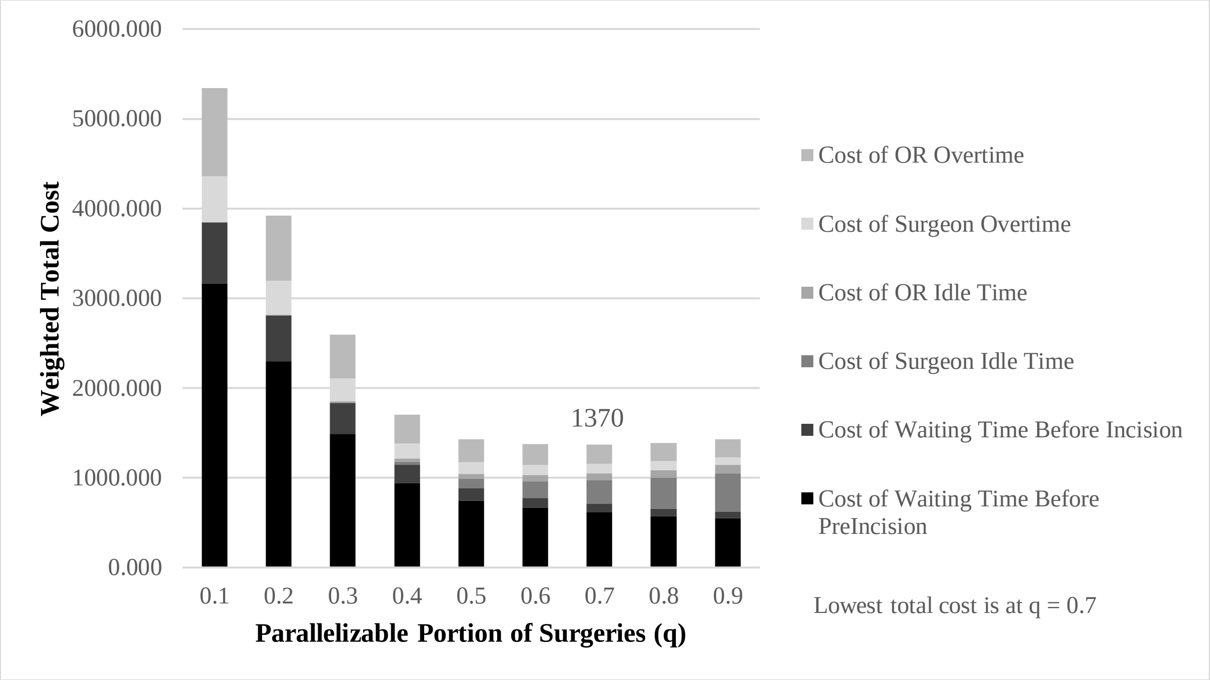

Figure 14. The breakdown of total cost under policy 1 and near-optimal schedule of Cost Scenario 1 for varying levels of the parallelizable portion of surgeries

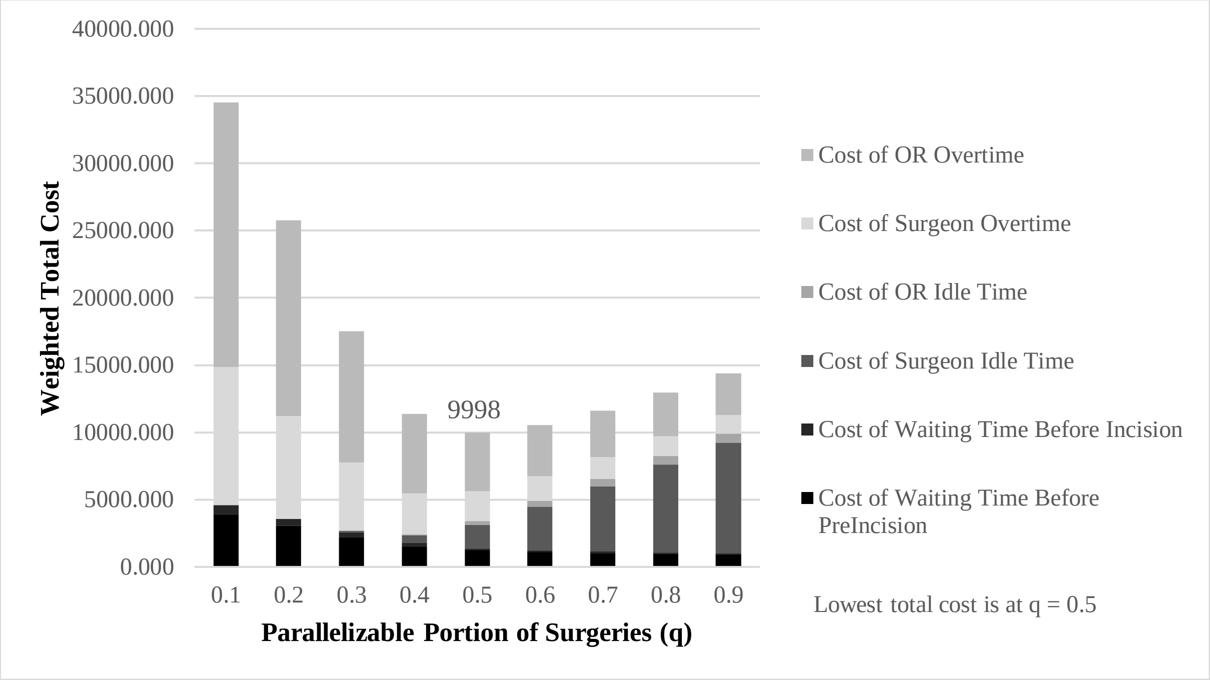

Figure 15. The breakdown of total cost under policy 1 and near-optimal schedule of Cost Scenario 2 for varying levels of the parallelizable portion of surgeries

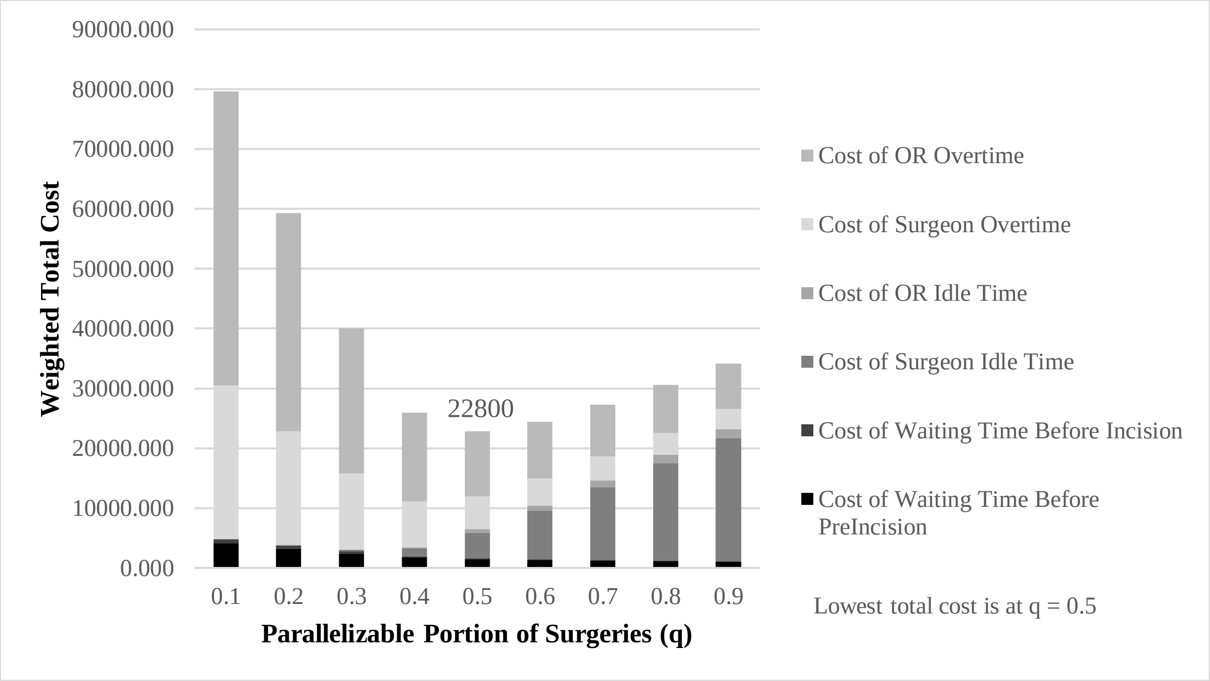

Figure 16. The breakdown of total cost under policy 1 and near-optimal schedule of Cost Scenario 3 for varying levels of the parallelizable portion of surgeries

For greater q values, lower waiting times are observed, and this decreases the total cost. The lowest total cost values are observed for q values of 0.7, 0.5, and 0.5 as 1370, 9998, and 22800 in the three cost configurations respectively. However, as q increases, surgeon is required for a smaller portion of the surgery duration and remains idle more frequently. This is seen as higher surgeon idle time costs at higher q values. This increase begins to overcome the lowering cost of waiting time and increases total cost for q values greater than 0.5 and 0.7 in the three cost configurations respectively. This finding indicates that, unlike OR cost, the total cost does not continuously decrease with increasing parallelizable portion of surgeries.

4.2.2 Policy 2: 2 Surgeons Using 3 Operating Rooms

While opening two ORs for each surgeon can double their patient throughput, this may not be feasible under circumstances where the hospital does not have a sufficient number of available ORs for this policy. The second policy investigated in this thesis applies to the case where two surgeons performing 12 surgeries in their ORs for the entire session (480 minutes) sharing half the session duration (240 minutes) of a third OR. This policy assumes that one surgeon performs six surgeries, then the other surgeon performs another six surgeries in the third OR. Similar to the policy 1, surgeries performed in the shared OR are scheduled with a 19 minutes shift from the near-optimal schedule and the other two ORs are scheduled using the near-optimal schedules found in section 4.1 according to the cost coefficient set being tested. A deterministic example of the scenarios tested for this case is illustrated on Figure 17. The numbers on the scheduled surgeries indicate the surgeon who performs them.

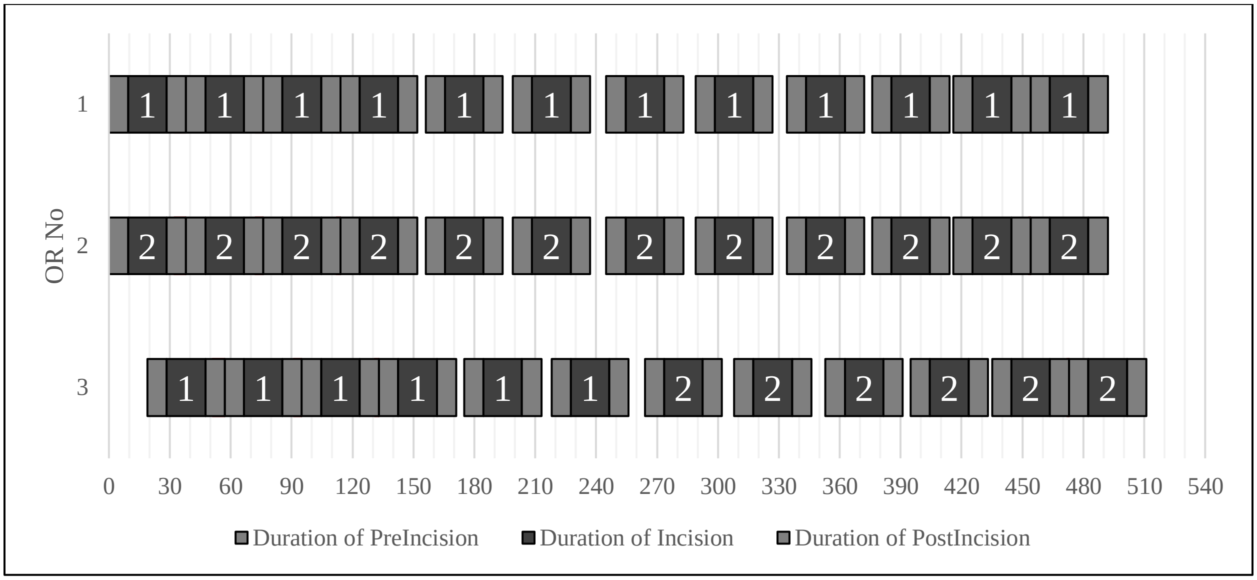

Figure 17. A deterministic example of Policy 2 scenarios

This scheduling policy is tested for the three cost configurations shown in Table 4. Same as the simulations performed in section 4.2.1, nine scenarios are simulated with 5000 replications of Monte Carlo simulations for the different values of the parallelizable portion of surgeries from 0.1 to 0.9. The resulting mean values and 95% confidence intervals for each of the cost components are reported in Appendix D. The expected value of the total weighted OR cost is calculated with the expression (38) using the respective cost coefficient sets for each of the simulation scenarios as described in Table 4. The expected total weighted OR cost values found for $q = 0.1, \ldots, 0.9$ are shown on Figure 18, Figure 19, and Figure 20 for cost configurations 1,2, and 3, respectively.

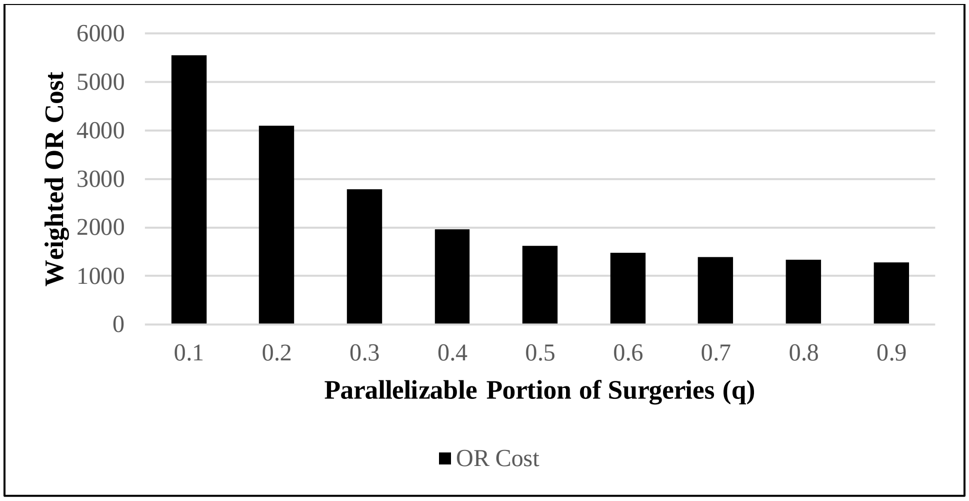

Figure 18. Total weighted OR cost for Cost Scenario 1

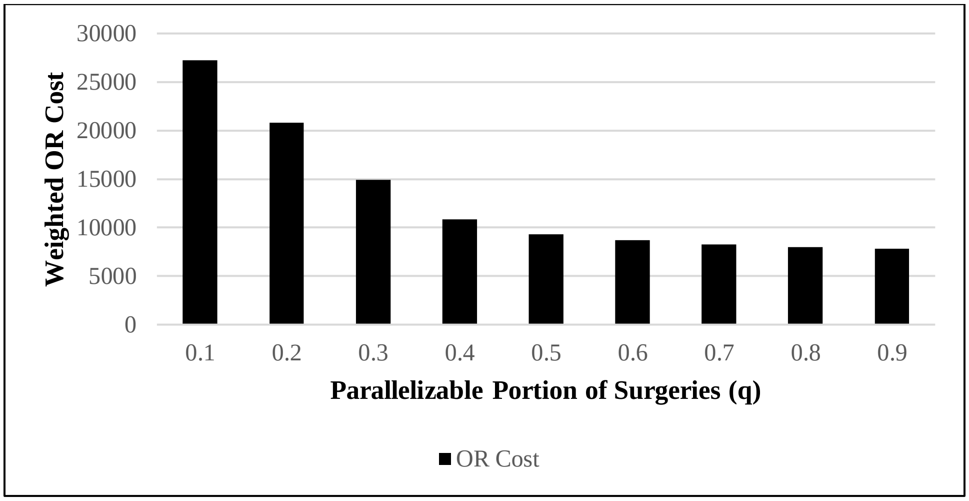

Figure 19. Total weighted OR cost for Cost Scenario 2

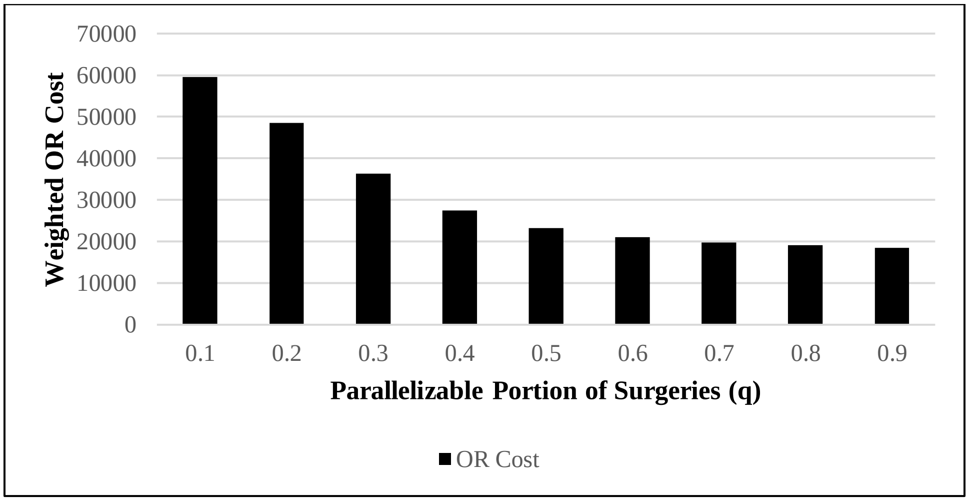

Figure 20. Total weighted OR cost for Cost Scenario 3

The simulation results of this scheduling policy show a similar pattern to the ones observed in policy 1 results in section 4.2.1. Again, a decline in OR cost is observed for the surgeries with higher parallelizable portions. For q values less than 0.5, the disruption described in section 4.2.1 is observed as greater OR costs for lower q values. These patterns are consistent under the three cost configurations tested and they are in line with the earlier findings in the literature.

In the next step of analysis, surgeon idle time $I_i^S(\omega)$ and overtime $O_m^S(\omega)$ are taken into account and the breakdown of the total cost calculated with the Expression (39) is observed and illustrated on Figure 21, Figure 22, and Figure 23 for cost configurations 1,2, and 3, respectively.

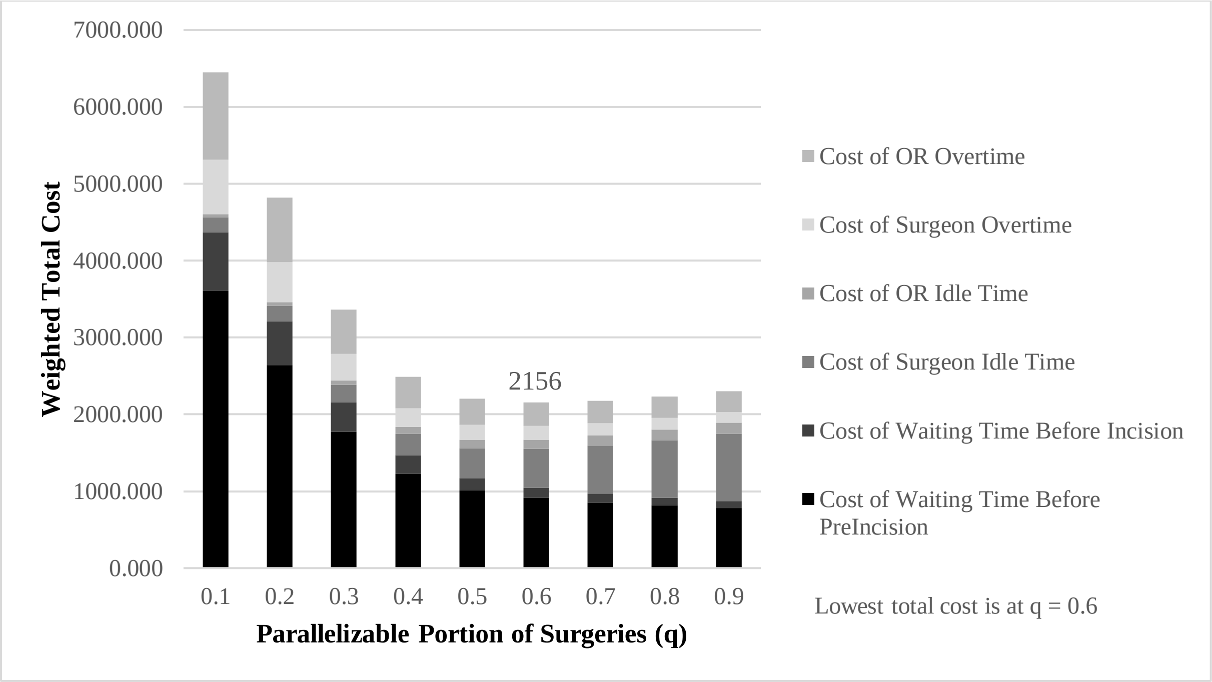

Figure 21. The breakdown of total cost under Policy 2 and near-optimal schedule of Cost Scenario 2 for varying levels of the parallelizable portion of surgeries

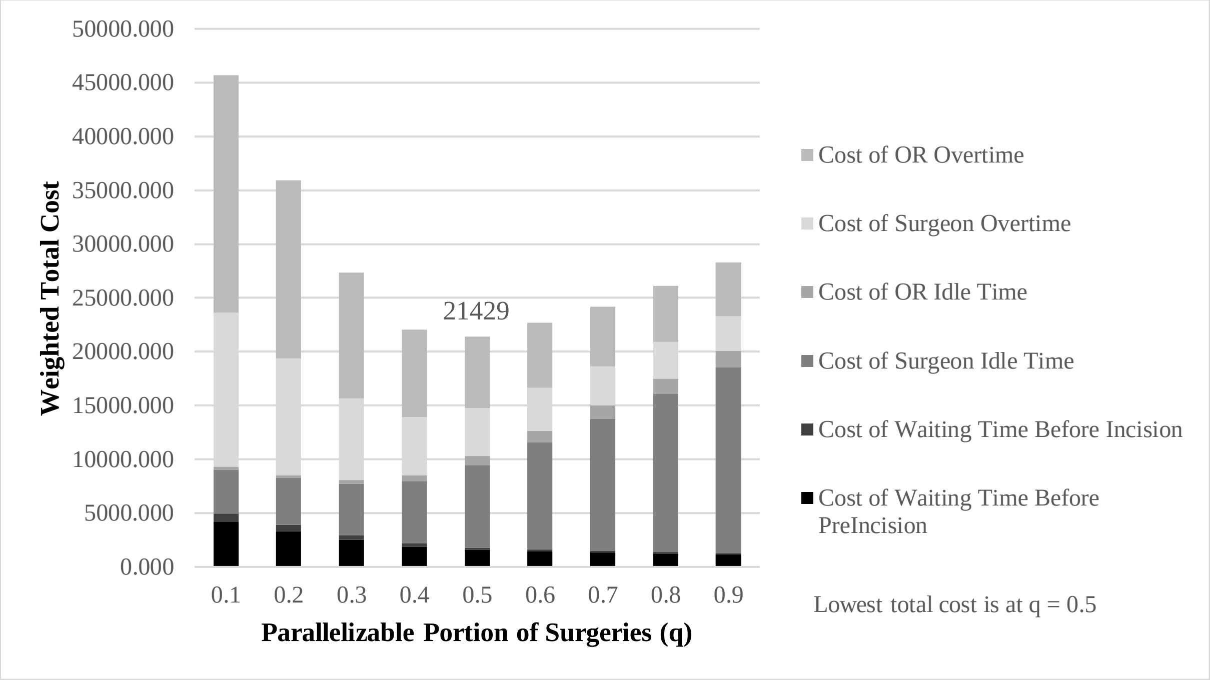

Figure 22. The breakdown of total cost under Policy 2 and near-optimal schedule of Cost Scenario 2 for varying levels of the parallelizable portion of surgeries

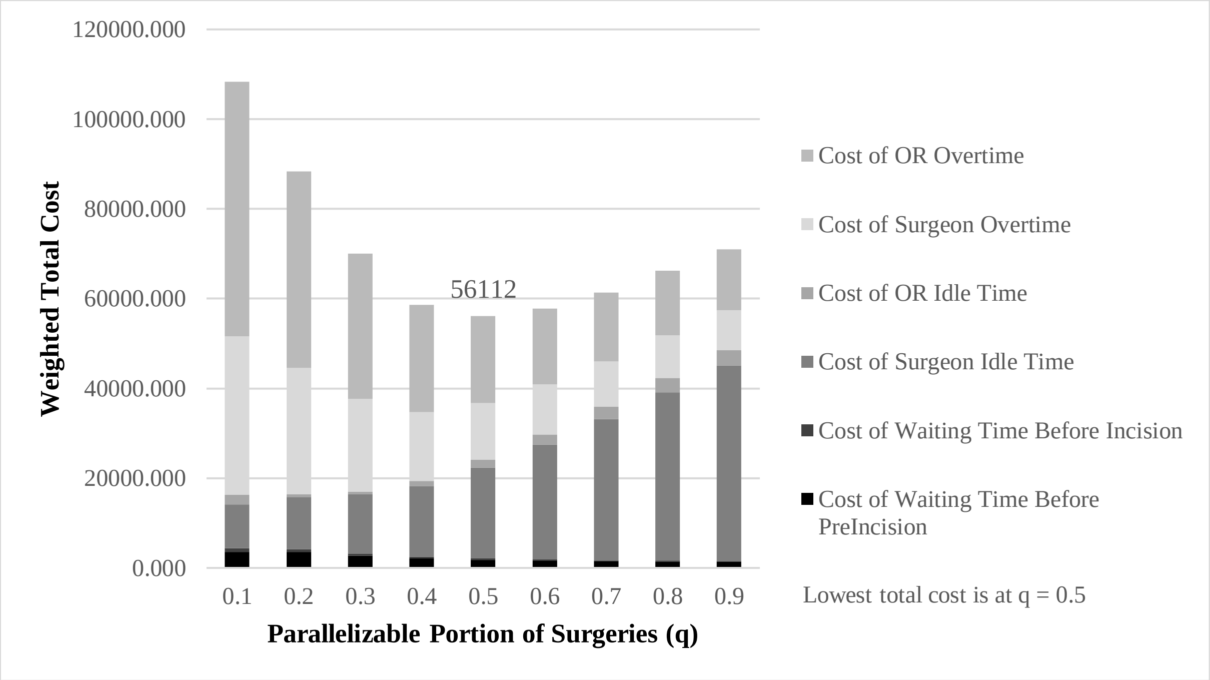

Figure 23. The breakdown of total cost under Policy 2 and near-optimal schedule of Cost Scenario 3 for varying levels of the parallelizable portion of surgeries

For surgeries with greater parallelizable portions, this policy yields lower waiting times. This decreases the total cost until the increase in surgeon idle time begins affecting the total cost. The lowest total cost values are observed for q values of 0.5, 0.5, and 0.6 as 2156, 21429, and 56112 in the three cost configurations respectively. This finding indicates that surgeries which have 50% to 60% of their duration comprised of non-critical activities are the most suitable to be scheduled using this policy.

5 Conclusions and Managerial Implications

In this thesis, cost implications of parallel surgery scheduling policies are investigated. First, three near-optimal appointment schedules are obtained for total cost objectives with differently weighted cost components using a simulation optimization method. Then, two parallel surgery scheduling policies are tested using simulating cost scenarios and the near-optimal appointment schedules obtained in the first step. For a setting where each surgeon works in one OR for the entire surgery duration, the cost outcomes of opening additional ORs and having the surgeon attend the critical portions of the surgeries in both ORs are tested under these policies. Since prior work in the literature has shown that OR to surgeon ratios beyond 2:1 lead to low utilization of the expensive resources (Berg et al. 2010), the two policies tested in this thesis are designed to have OR to surgeon ratios of 2:1 in Policy 1 and 1.5:1 in Policy 2. For both policies, OR costs (waiting time, OR idle time, OR overtime) and the breakdown of the total cost (surgeon idle time and overtime in addition to OR costs) are obtained under varying levels of parallelizable portion of surgeries and different sets of cost coefficients.

Across the different cost configurations and the two policies tested, it is consistently found that as the parallelizable portion of surgeries increases, parallel scheduling policies become more favorable when OR costs are the only consideration. This result is consistent with the prior findings in the literature. However, this analysis does not consider the costs related to surgeons' time which is another expensive resource.

Table 5

Parallelizable portion of surgery types that result in lowest total cost

Cost Coefficients			Parallelizable Portion (q) of the Surgery Type That Result in Lowest Total Cost	
c_W	c_I	c_O	Under Policy 1	Under Policy 2
1	1	1.5	70%	60%
1	20	30	50%	50%
1	50	75	50%	50%

Further analysis revealed that when surgeon idle time and surgeon overtime costs are taken into account, total cost does not continuously decrease with higher parallelizable portion of surgeries as suggested in the prior work. It is found that when the critical portion of surgeries comprise a lower portion of the total surgery duration, surgeons remain idle for longer periods during the session. This results in higher surgeon idle time costs for surgeries with high parallelizable portions. For surgeries which have less than 50% to 70% (based on the relative costs of waiting time, idle time, and overtime as shown in Table 5) of their durations parallelizable, this increase is justified by the decrease in overtime and waiting time costs. For surgeries with parallelizable portions over these q values, however, opening additional ORs becomes more costly under both policies because the surgeon idle time cost increases the total cost despite the decrease in overtime and waiting time. This increase in total cost is more pronounced in scenarios with cost coefficient sets 2 or 3 where the total cost is calculated using greater weights for Idle Time and Overtime relative to Waiting Time.

In many hospitals, the relative cost of idle time and overtime is much greater than the cost of patient waiting time, similar to the cost configurations 2 and 3. Based on the results of these scenarios, the surgery types that achieve the lowest cost are the ones that have 50% of their duration parallelizable. This finding is consistent in both policies as shown on Table 5. Therefore, it is concluded that surgeries which have 50% of their durations comprised of non-critical activities where the surgeon is not required are most suitable for parallel surgery scheduling policies, depending on the relative cost of idle time and overtime. In addition, it is found that surgeries with parallelizable portions less than 50% are not well suited for the tested parallel surgery scheduling policies as they result in much higher total costs.

This thesis contributes to the literature in several ways. First, it provides novel findings about the parallel surgery scheduling practices. The health outcomes of this practice are widely studied in the literature, but economic implications were only considered by one study (Batun et al. 2011). This thesis demonstrates how the weighted cost of patient waiting time, resource idle time, and overtime changes with the varying levels of parallelizable portion of surgeries under different cost scenarios and parallel surgery processing policies. It provides unique insights by investigating the resource idle time and overtime separately for surgeons and ORs. Second, this thesis contributes to the literature by providing more evidence for the efficacy of the dome rule with the near-optimal appointment times for uniform appointments given under different cost scenarios. Third, this study is the first to report experiment results of parallel surgery scheduling policies using real data for surgery durations to provide realistic insights. Lastly, this thesis provides further evidence for the robustness of simulation optimization methods in addressing appointment scheduling problems.

A limitation of this thesis is that it cannot point to particular surgery types as the most suitable ones for parallel scheduling. This is because the parallelizable portion of surgery is not standard for a surgery type across all hospitals for all surgical teams. Therefore, the managers of hospitals should identify the parallelizable portions of different surgery types for the conditions of their organization to benefit from the findings of this study. Another limitation of this thesis is the assumption of all surgeries scheduled in an OR being uniform in terms of surgery durations and the parallelizable portions. Further research is needed to investigate the cost implications of parallel surgery scheduling policies under conditions where surgeries with a mix of these attributes are considered.

References

Arena (Version 10.0.00) [Computer software]. Rockwell Software, Inc.

Bailey, N. T. J. (1952). A Study of Queues and Appointment Systems in Hospital Out-Patient Departments, with Special Reference to Waiting-Times. *Journal of the Royal Statistical Society. Series B (Methodological)*, *14*(2), 185–199.

Bam, M., Denton, B. T., Oyen, M. P. V., & Cowen, M. E. (2017). Surgery scheduling with recovery resources. *IISE Transactions*, *49*(10), 942–955. https://doi.org/10.1080/24725854.2017.1325027

Barnoon, S., & Wolfe, H. (1968). Scheduling A Multiple Operating Room System. *Health Services Research*, *3*(4), 272–285.

Batun, S., Denton, B. T., Huschka, T. R., & Schaefer, A. J. (2011). Operating Room Pooling and Parallel Surgery Processing Under Uncertainty. *INFORMS Journal on Computing*, *23*(2), 220–237. https://doi.org/10.1287/ijoc.1100.0396

Beasley, G. M., Pappas, T. N., & Kirk, A. D. (2015). Procedure Delegation by Attending Surgeons Performing Concurrent Operations in Academic Medical Centers: Balancing Safety and Efficiency. *Annals of Surgery*, *261*(6), 1044–1045. https://doi.org/10.1097/SLA.0000000000001208

Berg, B., Denton, B., Nelson, H., Balasubramanian, H., Rahman, A., Bailey, A., & Lindor, K. (2010). A discrete event simulation model to evaluate operational performance of a colonoscopy suite. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, *30*(3), 380–387. https://doi.org/10.1177/0272989X09345890

Blake, J. T., & Donald, J. (2002). Mount Sinai Hospital Uses Integer Programming to Allocate Operating Room Time. *Interfaces*, *32*(2), 63–73. https://doi.org/10.1287/inte.32.2.63.57

Boodman, S. G. (2017, July 10). Is your surgeon double-booked? *The Washington Post.* Retrieved from https://www.washingtonpost.com/national/health-science/is-your-surgeon-double-booked/2017/07/10/64a753f0-3a7d-11e7-9e48-c4f199710b69_story.html

Canadian Institute for Health Information. (2017). *Data Quality Documentation, Discharge Abstract Database—Current-Year Information, 2016–2017* (p. 18).

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, *119*(2), 354–366. https://doi.org/10.1016/j.ijpe.2009.03.009

Cayirli, T., & Veral, E. (2003). Outpatient Scheduling in Health Care: A Review of Literature*. *Production and Operations Management; Muncie*, *12*(4), 519–549.

Cayirli, T., Veral, E., & Rosen, H. (2006). Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science*, *9*(1), 47–58. https://doi.org/10.1007/s10729-006-6279-5

Cheng, T. C. E., & Sin, C. C. S. (1990). A state-of-the-art review of parallel-machine scheduling research. *European Journal of Operational Research*, *47*(3), 271–292. https://doi.org/10.1016/0377-2217(90)90215-W

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

Denton, B., Rahman, A., Nelson, H., & Bailey, A. (2006). Simulation of a Multiple Operating Room Surgical Suite (pp. 414–424). IEEE. https://doi.org/10.1109/WSC.2006.323110

Denton, B. T., Miller, A. J., Balasubramanian, H. J., & Huschka, T. R. (2010). Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty. *Operations Research*, *58*(4), 802-816,1028-1031. https://doi.org/10.1287/opre.1090.0791

Denton, Viapiano, J., & Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, *10*(1), 13–24. https://doi.org/10.1007/s10729-006-9005-4

Dexter, F., Macario, A., & Traub, R. D. (1999). Which Algorithm for Scheduling Add-on Elective Cases Maximizes Operating Room Utilization?: Use of Bin Packing Algorithms and Fuzzy Constraints in Operating Room Management. *Anesthesiology*, *91*(5), 1491. https://doi.org/10.1097/00000542-199911000-00043

Erdogan, S. A., & Denton, B. T. (2011). Surgery Planning and Scheduling. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, & J. C. Smith, *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved from http://doi.wiley.com/10.1002/9780470400531.eorms0861

Fei, H., Meskens, N., & Chu, C. (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, *58*(2), 221–230. https://doi.org/10.1016/j.cie.2009.02.012

Ferris, M. C., Deng, J. W. G., Fryback, D. G., & Kuruchittham, V. (2005). Breast cancer epidemiology: Calibrating simulations via optimization. *Oberwolfach Reports*, *2*.

Fetter, R. B., & Thompson, J. D. (1965). The Simulation of Hospital Systems. *Operations Research*, *13*(5), 689–711. https://doi.org/10.1287/opre.13.5.689

Fu, M. C. (2002). Optimization for simulation: Theory vs. Practice. *INFORMS Journal on Computing*, *14*(3), 192–215. https://doi.org/10.1287/ijoc.14.3.192.113

Gerchak, Y., Gupta, D., & Henig, M. (1996). Reservation Planning for Elective Surgery under Uncertain Demand for Emergency Surgery. *Management Science*, (3), 321.

Gilmore, P. C., & Gomory, R. E. (1964). Sequencing a One State-Variable Machine: A Solvable Case of the Traveling Salesman Problem. *Operations Research*, *12*(5), 655–679. https://doi.org/10.1287/opre.12.5.655

Glover, F. (1977). Heuristics for Integer Programming Using Surrogate Constraints. *Decision Sciences*, *8*(1), 156–166. https://doi.org/10.1111/j.1540-5915.1977.tb01074.x

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, *13*(5), 533–549. https://doi.org/10.1016/0305-0548(86)90048-1

Glover, F., Kelly, J. P., & Laguna, M. (1996). New Advances and Applications of Combining Simulation and Optimization. In *Proceedings of the 28th Conference on Winter Simulation* (pp. 144–152). Washington, DC, USA: IEEE Computer Society. https://doi.org/10.1145/256562.256595

Gosavi, A. (2015). *Simulation-Based Optimization Parametric Optimization Techniques and Reinforcement Learning*. New York: Springer.

Guan, J., Karsy, M., Brock, A. A., Couldwell, W. T., & Schmidt, R. H. (2017). Overlapping Surgery: A Review of the Controversy, the Evidence, and Future Directions. *Neurosurgery*, *64*(CN_suppl_1), 110–113. https://doi.org/10.1093/neuros/nyx200

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, *14*(1), 89–114. https://doi.org/10.1007/s10729-010-9143-6

Guinet, A., & Chaabane, S. (2003). Operating theatre planning. *International Journal of Production Economics*, *85*(1), 69–81. https://doi.org/10.1016/S0925-5273(03)00087-2

Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, *40*(9), 800–819. https://doi.org/10.1080/07408170802165880

Hans, E., Wullink, G., van Houdenhoven, M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*(3), 1038–1050. https://doi.org/10.1016/j.ejor.2006.08.022

Hanss, R., Buttgereit, B., Tonner, P. H., Bein, B., Schleppers, A., Steinfath, M., … Bauer, M. (2005). Overlapping Induction of Anesthesia—An Analysis of Benefits and Costs. *The Journal of the American Society of Anesthesiologists*, *103*(2), 391–400.

Ho, C.-J., & Lau, H.-S. (1992). Minimizing Total Cost in Scheduling Outpatient Appointments. *Management Science*, *38*(12), 1750–1764.

Irizarry, M. D. L. A., Wilson, J. R., & Trevino, J. (2001). A Flexible Simulation Tool for Manufacturing-cell Design. *IIE Transactions*, *33*(10), 837–846. https://doi.org/10.1023/A:1010970504862

Jebali, A., & Diabat, A. (2017). A Chance-constrained operating room planning with elective and emergency cases under downstream capacity constraints. *Computers & Industrial Engineering*, *114*, 329–344. https://doi.org/10.1016/j.cie.2017.07.015

Jebali, A., Hadj Alouane, A. B., & Ladet, P. (2006). Operating rooms scheduling. *International Journal of Production Economics*, *99*(1), 52–62. https://doi.org/10.1016/j.ijpe.2004.12.006

Jiang, B., Tang, J., & Yan, C. (2019). A stochastic programming model for outpatient appointment scheduling considering unpunctuality. *Omega*, *82*, 70–82. https://doi.org/10.1016/j.omega.2017.12.004

Jun, J. B., Jacobson, S. H., & Swisher, J. R. (1999). Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society*, *50*(2), 109–123. https://doi.org/10.1057/palgrave.jors.2600669

Klassen, K. J., & Yoogalingam, R. (2008). An assessment of the interruption level of doctors in outpatient appointment scheduling. *Operations Management Research*, *1*(2), 95–102. https://doi.org/10.1007/s12063-008-0013-z

Klassen, K. J., & Yoogalingam, R. (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, *18*(4), 447–458.

Kleijnen, J. P. C., & Wan, J. (2007). Optimization of simulated systems: OptQuest and alternatives. *Simulation Modelling Practice and Theory*, *15*(3), 354–362. https://doi.org/10.1016/j.simpat.2006.11.001

Laguna, M. (1997a). Metaheuristic optimization with evolver, genocop and optquest. In *EURO/INFORMS Joint International Meeting, Plenaries and Tutorials* (pp. 141–150).

Laguna, M. (1997b). Optimization of complex systems with OptQuest. *A White Paper from OptTek Systems, Inc*.

Laguna, M., & Wubbena, T. (2005). Modeling and Solving a Selection and Assignment Problem. In B. Golden, S. Raghavan, & E. Wasil (Eds.), *The Next Wave in Computing, Optimization, and Decision Technologies* (pp. 149–162). Springer US.

Lamiri, M., Grimaud, F., & Xie, X. (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics*, *120*(2), 400–410. https://doi.org/10.1016/j.ijpe.2008.11.021

Lamiri, M., Xie, X., Dolgui, A., & Grimaud, F. (2008). A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, *185*(3), 1026–1037. https://doi.org/10.1016/j.ejor.2006.02.057

Lamiri, M., Xie, X., & Zhang, S. (2008). Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions*, *40*(9), 838–852. https://doi.org/10.1080/07408170802165831

Law, A. M., & Kelton, W. D. (1991). *Simulation modeling and analysis* (2nd ed). New york: McGraw-Hill.

Lin, R.-C., Sir, M. Y., & Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega*, *41*(5), 881–892. https://doi.org/10.1016/j.omega.2012.11.003

Liu, J. B., Ban, K. A., Berian, J. R., Hutter, M. M., Huffman, K. M., Liu, Y., … Ko, C. Y. (2017). Concurrent bariatric operations and association with perioperative outcomes: Registry based cohort study. *BMJ*, *358*, j4244. https://doi.org/10.1136/bmj.j4244

Macario, A., Vitez, T., Dunn, B., & McDonald, T. (1995). Where Are the Costs in Perioperative Care?: Analysis of Hospital Costs and Charges for Inpatient Surgical Care. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, *83*(6), 1138–1144.

Magerlein, J. M., & Martin, J. B. (1978). Surgical demand scheduling: A review. *Health Services Research*, *13*(4), 418–433.

May, J. H., Strum, D. P., & Vargas, L. G. (2000). Fitting the Lognormal Distribution to Surgical Procedure Times*. *Decision Sciences*, *31*(1), 129–148. https://doi.org/10.1111/j.1540-5915.2000.tb00927.x

OECD (Ed.). (2004). *Towards high-performing health systems: Policy studies*. Paris: OECD.

OECD. (2013). What Future for Health Spending? *OECD Economics Department Policy Notes*, No. 19.

OECD. (2017). *Health at a Glance 2017*. https://doi.org/10.1787/health_glance-2017-en

OptQuest (Version 5.1.1.2) [Computer software]. OptTek Systems, Inc.

Palisade Corporation. (2002). Risk analysis and Simulation add-in for Microsoft Excel. *Pallisade Corporation, New York*.

Patterson, P. (1996). What makes a well-oiled scheduling system? *OR Manager*, *12*(9), 19–23.

Ramasesh, R. (1990). Dynamic job shop scheduling: A survey of simulation research. *Omega*, *18*(1), 43–57. https://doi.org/10.1016/0305-0483(90)90017-4

Ravi, B., Pincus, D., Wasserstein, D., Govindarajan, A., Huang, A., Austin, P. C., … Kreder, H. J. (2018). Association of Overlapping Surgery With Increased Risk for Complications Following Hip Surgery: A Population-Based, Matched Cohort Study. *JAMA Internal Medicine*, *178*(1), 75–83. https://doi.org/10.1001/jamainternmed.2017.6835

Richards, J., Brown, A., & Homan, C. (2002). *The data quality study of the Canadian Discharge Abstract Database*. Citeseer.

Roberts, R. R. (1999). Distribution of Variable vs Fixed Costs of Hospital Care. *JAMA*, *281*(7), 644. https://doi.org/10.1001/jama.281.7.644

Robinson, L. W., & Chen, R. R. (2003). Scheduling doctors' appointments: Optimal and empirically-based heuristic policies. *IIE Transactions*, *35*(3), 295–307. https://doi.org/10.1080/07408170304367

Rockwell Automation. (2012). OptQuest for Arena User's Guide.

Santibáñez, P., Begen, M., & Atkins, D. (2007). Surgical block scheduling in a system of hospitals: An application to resource and wait list management in a British Columbia health authority. *Health Care Management Science*, *10*(3), 269–282. https://doi.org/10.1007/s10729-007-9019-6

Saremi, A., Jula, P., ElMekkawy, T., & Wang, G. G. (2013). Appointment scheduling of outpatient surgical services in a multistage operating room department. *International Journal of Production Economics*, *141*(2), 646–658. https://doi.org/10.1016/j.ijpe.2012.10.004

Testi, A., Tanfani, E., & Torre, G. (2007). A three-phase approach for operating theatre schedules. *Health Care Management Science*, *10*(2), 163–172. https://doi.org/10.1007/s10729-007-9011-1

Torkki, P. M., Marjamaa, R. A., Torkki, M. I., Kallio, P. E., & Kirvelä, O. A. (2005). Use of Anesthesia Induction Rooms Can Increase the Number of Urgent Orthopedic Cases Completed within 7 Hours. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, *103*(2), 401–405.

Tsai, S. C., & Chen, S. T. (2017). A simulation-based multi-objective optimization framework: A case study on inventory management. *Omega*, *70*, 148–159. https://doi.org/10.1016/j.omega.2016.09.007

Wang, P. P. (1993). Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics (NRL)*, *40*(3), 345–360. https://doi.org/10.1002/1520-6750(199304)40:3<345::AID-NAV3220400305>3.0.CO;2-N

Weiss, E. N. (1990). Models for Determining Estimated Start Times and Case Orderings In Hospital Operating Rooms-. *IIE Transactions*, *22*(2), 143–150. https://doi.org/10.1080/07408179008964166

Wullink, G., Van Houdenhoven, M., Hans, E. W., van Oostrum, J. M., van der Lans, M., & Kazemier, G. (2007). Closing Emergency Operating Rooms Improves Efficiency. *Journal of Medical Systems*, *31*(6), 543–546. https://doi.org/10.1007/s10916-007-9096-6

Zhang, A. L., Sing, D. C., Dang, D. Y., Ma, C. B., Black, D., Vail, T. P., & Feeley, B. T. (2016). Overlapping Surgery in the Ambulatory Orthopaedic Setting. *The Journal of Bone and Joint Surgery. American Volume*, *98*(22), 1859–1867. https://doi.org/10.2106/JBJS.16.00248

Zhang, B., Murali, P., Dessouky, M. M., & Belson, D. (2009). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, *60*(5), 663–673. https://doi.org/10.1057/palgrave.jors.2602596

Zhang, Z., & Xie, X. (2015). Simulation-based optimization for surgery appointment scheduling of multiple operating rooms. *IIE Transactions*, *47*(9), 998–1012. https://doi.org/10.1080/0740817X.2014.999900

Appendix A. Data Elements Received from CIHI

Table A1

Descriptions of the Data Elements Received from CIHI

Data Element	Description
SUBMITTING_PROV_CODE	Deidentified province code (First character of the Institution Code)
DEID_INST_CODE	A deidentified five-character code assigned to a reporting facility by a provincial/ territorial ministry of health identifying the facility and the level of care of the data submitted
FISCAL_YEAR	Same in every row: 2016
ADMISSION_DATE	The date and time that the patient was officially registered as an inpatient.
ADMISSION_TIME	The time that the patient was officially registered as an inpatient.
ADMIT_BY_AMBULANCE_IND	Identifies whether a patient arrives at the health care facility via ambulance and the type of ambulance that was used.
DISCHARGE_DATE	The date when the patient was formally discharged.
DISCHARGE_TIME	The time when the patient was formally discharged.
DISCHARGE_DISPOSITION	The location (01 to 05) where the patient was discharged to or the status of the patient on discharge (06 to 09 and 12). 01 Discharged Home (private dwelling, not an institution; no support services) 02 Patient left at his/her own risk following registration. Triage (if an ED visit), further assessment by a service provider and treatment did not occur. 03 Patient left the emergency department at his/her own risk following registration and triage. Further assessment by a service provider and treatment did not occur. Note: this visit disposition is limited to ED visits 04 Patient left at his/her own risk following registration, triage (if an ED Visit), and further assessment by a service provider. Treatment did not occur.

Data Element	Description
	05 Patient left at his/her own risk following registration, triage (if an ED Visit), further assessment by a service provider and initiation of treatment
DIAG_CODE_1 - 25	The ICD-10-CA classification code that describes the diagnoses, conditions, problems or circumstances of the patient during the length of stay in the health care facility.
INTERV_EPISODE_START_DATE_1 - 20	The date when the patient enters a physical area (intervention location) to have a service(s) (intervention) initiated.
INTERV_EPISODE_START_TIME_1 - 20	The time when the patient enters a physical area (intervention location) to have a service(s) (intervention) initiated.
INTERV_EPISODE_END_DATE_1 - 20	The date when the patient exits the physical area (Intervention Location) after service(s) ended.
INTERV_EPISODE_END_TIME_1 - 20	The time when the patient exits the physical area (Intervention Location) after service(s) ended.
INTERV_CCI_CODE_1 - 20	A valid CCI code(s) describing the services (procedures/intervention) performed for or on behalf of the patient to improve health.
DEID_INTERV_PROVIDER_NUM_1 - 20	A unique deidentified identifier of the health care providers (physicians and allied health care professionals) involved in each intervention.
INTERV_LOCATION_NUM_1 - 20	Records the physical area in the health care facility where a service(s) (intervention) took place.
DEID_PROVIDER_NUM_1 - 10	A deidentified identifier number associated with the provider responsible for provision of services to the patient during the visit.
RECORD_ID	Unique Record Identifier

Appendix B. A Deterministic Example for Case 1 Scenarios

Table B2
Values of the parameters for a deterministic example (Columns 1-13)

i	b_i	h_i^S	y_i	h_i^R	u_i^S	u_i^R	v_i^S	v_i^R	x_i	$W_i^R(\omega)$	$s_i^{pre}(\omega)$	$t_i(\omega)$
1	1	1	1	1	First	First	13	2	0	0	0	38
2	1	3	1	2	13	1	14	3	32	6	38	38
3	1	5	1	3	14	2	15	4	72	4	76	38
4	1	7	1	4	15	3	16	5	112	2	114	38
5	1	9	1	5	16	4	17	6	156	0	156	38
6	1	11	1	6	17	5	18	7	199	0	199	38
7	1	13	1	7	18	6	19	8	245	0	245	38
8	1	15	1	8	19	7	20	9	289	0	289	38
9	1	17	1	9	20	8	21	10	334	0	334	38
10	1	19	1	10	21	9	22	11	376	0	376	38
11	1	21	1	11	22	10	23	12	416	0	416	38
12	1	23	1	12	23	11	24	Last	453	1	454	38
13	1	2	2	1	1	Last	2	14	19	0	19	38
14	1	4	2	2	2	13	3	15	51	6	57	38
15	1	6	2	3	3	14	4	16	91	4	95	38
16	1	8	2	4	4	15	5	17	131	2	133	38
17	1	10	2	5	5	16	6	18	175	0	175	38
18	1	12	2	6	6	17	7	19	218	0	218	38
19	1	14	2	7	7	18	8	20	264	0	264	38
20	1	16	2	8	8	19	9	21	308	0	308	38
21	1	18	2	9	9	20	10	22	353	0	353	38
22	1	20	2	10	10	21	11	23	395	0	395	38
23	1	22	2	11	11	22	12	24	435	0	435	38
24	1	24	2	12	12	23	Last	Last	472	1	473	38

Table B3

Values of the parameters for a deterministic example (Columns 14-26)

$pre_i(\omega)$	$s_i^{pre}(\omega) + pre_i(\omega)$	$W_i^S(\omega)$	$s_i^p(\omega)$	$p_i(\omega)$	$s_i^p(\omega) + p_i(\omega)$	$s_i^{post}(\omega)$	$post_i(\omega)$	$s_i^{post}(\omega) + post_i(\omega)$	$I_i^S(\omega)$	$I_i^R(\omega)$	$O_m^S(\omega)$	$O_k^R(\omega)$
9.5	9.5	0	9.5	19	28.5	28.5	9.5	38	0	0	0	0
9.5	47.5	0	47.5	19	66.5	66.5	9.5	76	0	0	0	0
9.5	85.5	0	85.5	19	104.5	104.5	9.5	114	0	0	0	0
9.5	123.5	0	123.5	19	142.5	142.5	9.5	152	0	0	0	0
9.5	165.5	0	165.5	19	184.5	184.5	9.5	194	4	4	0	0
9.5	208.5	0	208.5	19	227.5	227.5	9.5	237	5	5	0	0
9.5	254.5	0	254.5	19	273.5	273.5	9.5	283	8	8	0	0
9.5	298.5	0	298.5	19	317.5	317.5	9.5	327	6	6	0	0
9.5	343.5	0	343.5	19	362.5	362.5	9.5	372	7	7	0	0
9.5	385.5	0	385.5	19	404.5	404.5	9.5	414	4	4	0	0
9.5	425.5	0	425.5	19	444.5	444.5	9.5	454	2	2	0	0
9.5	463.5	0	463.5	19	482.5	482.5	9.5	492	0	0	0	12
9.5	28.5	0	28.5	19	47.5	47.5	9.5	57	0	0	0	0
9.5	66.5	0	66.5	19	85.5	85.5	9.5	95	0	0	0	0
9.5	104.5	0	104.5	19	123.5	123.5	9.5	133	0	0	0	0
9.5	142.5	0	142.5	19	161.5	161.5	9.5	171	0	0	0	0
9.5	184.5	0	184.5	19	203.5	203.5	9.5	213	0	4	0	0
9.5	227.5	0	227.5	19	246.5	246.5	9.5	256	0	5	0	0
9.5	273.5	0	273.5	19	292.5	292.5	9.5	302	0	8	0	0
9.5	317.5	0	317.5	19	336.5	336.5	9.5	346	0	6	0	0
9.5	362.5	0	362.5	19	381.5	381.5	9.5	391	0	7	0	0
9.5	404.5	0	404.5	19	423.5	423.5	9.5	433	0	4	0	0
9.5	444.5	0	444.5	19	463.5	463.5	9.5	473	0	2	0	0
9.5	482.5	0	482.5	19	501.5	501.5	9.5	511	0	0	21.5	31

Appendix C. Extended Results of Policy 1 Simulations

Table C4

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 1 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 1$; $c_O = 1$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	3163.062 [1482.214, 5886.949]	680.847 [559.722, 864.336]	0.036 [0, 0.151]	3.217 [0, 15.356]	342.781 [205.248, 553.523]	655.115 [386.114, 1063.189]	4829.798 [2661.396, 8204.247]
0.2	2297.477 [824.875, 4644.329]	510.272 [412.861, 652.761]	0.779 [0, 7.163]	4.217 [0, 22.164]	254.095 [128.816, 435.961]	485.386 [237.469, 845.31]	3540.044 [1637.413, 6438.069]
0.3	1488.938 [207.429, 3796.793]	344.268 [240.01, 479.478]	6.125 [0, 33.114]	9.744 [0, 72.977]	170.138 [57.146, 345.934]	324.892 [101.466, 670.336]	2330.289 [678.284, 5169.574]
0.4	940.974 [44.469, 3220.042]	204.729 [80.12, 367.841]	35.625 [5.633, 90.997]	36.368 [0, 166.092]	110.451 [19.731, 281.587]	212.169 [32.184, 555.068]	1500.325 [338.083, 4349.072]
0.5	744.746 [24.761, 2932.13]	140.017 [30.176, 317.767]	101.391 [57.466, 173.301]	59.073 [0, 200.682]	86.811 [12.774, 264.834]	170.049 [22.683, 520.915]	1198.908 [283.726, 3919.513]
0.6	665.473 [20.681, 2798.862]	111.731 [11.763, 282.777]	180.84 [133.7, 266.143]	71.861 [0, 211.706]	77.006 [10.363, 244.235]	154.676 [21.908, 483.752]	1081.079 [265.549, 3706.893]
0.7	614.15 [21.732, 2776.175]	93.93 [3.118, 267.208]	262.621 [213.853, 365.387]	78.671 [0, 215.804]	69.616 [8.516, 235.354]	144.104 [21.527, 468.173]	1002.907 [258.099, 3644.177]
0.8	574.685 [19.201, 2676.071]	80.849 [0, 257.195]	346.389 [295.491, 464.409]	84.778 [0, 220.425]	63.992 [6.732, 223.402]	136.894 [20.812, 457.32]	945.653 [257.81, 3556.991]
0.9	549.541 [20.636, 2468.213]	70.367 [0, 237.411]	430.967 [378.169, 563.75]	89.888 [0, 220.349]	59.253 [5.49, 216.584]	131.482 [21.233, 444.343]	907.018 [253.348, 3318.648]

Table C5

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 1 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 20$; $c_O = 30$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	3907.008 [2228.293, 6680.219]	683.704 [563.989, 865.678]	0.038 [0, 0.312]	0.283 [0, 3.011]	342.64 [206.531, 546.965]	654.879 [385.159, 1062.226]	24242.73 [14526.82, 38921.42]
0.2	3041.139 [1532.694, 5450.589]	514.174 [421.835, 655.188]	0.766 [0, 6.791]	0.321 [0, 3.278]	254.139 [129.853, 435.901]	485.463 [242.017, 848.302]	18125.64 [9387.792, 31185.65]
0.3	2220.427 [860.424, 4500.902]	352.614 [279.765, 487.65]	5.595 [0, 32.483]	0.378 [0, 3.392]	169.679 [55.523, 349.119]	323.987 [96.964, 681.708]	12300.22 [4184.711, 25081.16]
0.4	1561.406 [221.116, 3903.396]	220.228 [132.6, 370.569]	27.044 [2.83, 81.978]	3.725 [0, 46.347]	102.269 [3.835, 285.145]	197.47 [17.297, 556.965]	7780.223 [1399.973, 20544.2]
0.5	1249.468 [97.721, 3641.255]	152.701 [47.214, 312.652]	85.903 [47.067, 158.873]	15.247 [0, 124.773]	73.109 [1.384, 246.223]	145.984 [11.986, 490.034]	6086.645 [1240.159, 18188.76]
0.6	1115.47 [83.458, 3356.623]	125.054 [25.832, 286.085]	162.161 [119.034, 249.498]	21.109 [0, 135.123]	61.225 [1.1, 230.41]	126.314 [9.467, 458.213]	5452.125 [1213.548, 17014.31]
0.7	1030.26 [72.349, 3292.703]	106.161 [11.558, 274.871]	242.738 [196.462, 352.739]	26.801 [0, 148.071]	54.071 [1.131, 225.013]	114.991 [7.865, 450.961]	5122.162 [1202.411, 16805.58]
0.8	980.389 [66.457, 3381.982]	93.315 [3.955, 269.763]	325.971 [277.613, 455.19]	31.602 [0, 156.779]	49.755 [1.193, 215.928]	108.341 [7.069, 443.15]	4955.968 [1224.632, 16531.38]
0.9	936.835 [63.427, 3167.084]	83.616 [0, 250.169]	410.152 [357.387, 555.742]	35.113 [0, 163.514]	46.026 [1.214, 205.471]	103.006 [6.683, 424.575]	4812.886 [1208.885, 16012.02]

Table C6

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 1 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 50$; $c_O = 75$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	4103.582 [2449.739, 6766.497]	684.128 [563.751, 863.643]	0.039 [0, 0.121]	0.012 [0, 0]	342.796 [205.069, 545.089]	655.203 [384.985, 1057.01]	53928.48 [32008.61, 86336.85]
0.2	3236.545 [1761.864, 5580.237]	514.535 [422.439, 655.117]	0.765 [0, 6.778]	0.012 [0, 0]	254.211 [131.817, 435.355]	485.599 [244.459, 850.276]	40171.64 [20541.37, 69397.22]
0.3	2411.682 [1059.903, 4767.131]	352.563 [281.587, 487.848]	5.411 [0, 32.23]	0.02 [0, 0.094]	169.424 [57.516, 349.182]	323.488 [102.292, 679.924]	27026.85 [9078.771, 55552.3]
0.4	1751.225 [408.971, 4061.589]	221.321 [137.547, 370.339]	27.067 [2.838, 81.79]	2.617 [0, 37.364]	102.323 [4.179, 281.748]	197.451 [17.055, 550.991]	16912.2 [2515.648, 45815.98]
0.5	1425.173 [192.523, 3860.975]	153.911 [56.484, 316.293]	85.387 [46.921, 159.679]	12.995 [0, 109.812]	72.395 [1.383, 245.969]	144.968 [12.195, 483.232]	13101.45 [2151.654, 40356.02]
0.6	1285.988 [151.84, 3638.451]	125.918 [30.61, 286.475]	161.615 [119.419, 252.443]	19.157 [0, 127.862]	60.703 [1.156, 235.12]	125.614 [9.317, 467.874]	11790.76 [2103.981, 38605.02]
0.7	1202.981 [133.912, 3524.354]	109.101 [16.378, 275.279]	242.724 [197.143, 351.522]	23.737 [0, 146.626]	53.838 [1.078, 220.528]	114.598 [7.989, 449.498]	11093.75 [2156.573, 37267.09]
0.8	1139.35 [123.251, 3409.173]	95.347 [6.75, 258.754]	325.319 [276.793, 449.088]	28.173 [0, 151.403]	48.948 [1.085, 211.99]	106.497 [6.678, 429.224]	10630.64 [2127.33, 35797.24]
0.9	1093.336 [117.692, 3317.591]	85.237 [1.423, 249.67]	409.247 [357.466, 548.279]	31.576 [0, 157.494]	45.224 [1.196, 200.338]	100.857 [7.002, 411.463]	10321.63 [2203.906, 33957.44]

Appendix D. Extended Results of Policy 2 Simulations

Table D7

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 2 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 1$; $c_O = 1$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	3601.972 [1557.97, 6882.503]	765.893 [567.352, 1079.934]	195.813 [63.897, 358.486]	43.635 [0, 98.154]	470.788 [238.33, 812.881]	756.496 [397.684, 1290.459]	5546.244 [2857.957, 9899.021]
0.2	2642.201 [825.015, 5519.741]	568.677 [410.963, 821.976]	203.001 [103.729, 353.621]	45.19 [0, 101.977]	345.242 [141.378, 649.975]	561.069 [242.884, 1038.754]	4097.671 [1719.71, 7887.677]
0.3	1775.552 [262.745, 4563.789]	382.756 [226.197, 620.118]	223.187 [146.453, 363.422]	55.029 [0, 152.313]	232.833 [69.007, 517.614]	384.977 [117.93, 833.802]	2790.804 [822.207, 6442.829]
0.4	1228.838 [104.373, 4023.502]	236.399 [85.027, 490.57]	283.629 [212.803, 435.991]	87.194 [1.28, 242.237]	161.575 [32.483, 438.235]	271.414 [54.809, 706.659]	1959.553 [522.455, 5516.833]
0.5	1007.231 [76.832, 3570.386]	165.734 [36.028, 407.51]	385.234 [311.359, 530.85]	110.846 [4.86, 275.106]	130.704 [21.119, 385.287]	224.1 [35.622, 621.12]	1619.961 [446.056, 4903.683]
0.6	912.042 [64.424, 3471.34]	133.903 [14.68, 377.809]	503.336 [427.75, 661.174]	123.563 [7.172, 291.623]	116.798 [18.17, 366.65]	205.088 [35.202, 596.104]	1477.14 [422.088, 4683.386]
0.7	853.525 [65.601, 3253.383]	113.712 [4.887, 345.07]	625.481 [549.836, 792.206]	131.589 [9.381, 296.926]	106.946 [15.996, 340.107]	193.08 [32.571, 584.76]	1388.445 [412.752, 4437.706]
0.8	816.504 [66.523, 3145.611]	98.769 [0.327, 329.716]	750.241 [670.814, 933.683]	138.605 [11.8, 300.716]	99.976 [14.361, 332.465]	185.368 [32.552, 566.737]	1331.93 [411.809, 4340.66]
0.9	782.525 [59.739, 3099.858]	86.749 [0, 313.619]	876.173 [792.145, 1075.536]	145.02 [11.809, 307.655]	94.085 [13.082, 316.276]	179.516 [31.44, 540.551]	1283.568 [406.02, 4228.728]

Table D8

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 2 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 20$; $c_O = 30$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	4191.686 [2169.359, 7406.629]	773.606 [559.105, 1097.559]	203.386 [62.758, 379.582]	13.189 [0, 55.851]	478.494 [238.435, 820.61]	733.908 [372.195, 1268.629]	27246.3 [14292.89, 46498.55]
0.2	3323.886 [1483.556, 6340.469]	591.834 [419.537, 867.41]	217.627 [109.864, 378.817]	13.685 [0, 55.159]	360.244 [150.701, 688.134]	552.995 [232.81, 1057.751]	20779.27 [9290.362, 38811.57]
0.3	2517.039 [775.218, 5366.098]	421.602 [285.307, 664.936]	240.094 [156.414, 402.145]	15.984 [0, 59.771]	253.999 [85.424, 564.689]	388.415 [111.526, 867.938]	14910.78 [5070.017, 32203.03]
0.4	1889.97 [340.831, 4772.775]	284.853 [149.111, 544.312]	290.1 [215.293, 463.394]	26.69 [0, 114.329]	180.37 [43.924, 470.807]	271.234 [60.67, 714.244]	10845.64 [3584.763, 26669.48]
0.5	1580.488 [226.788, 4373.14]	213.272 [79.819, 473.655]	384.408 [308.489, 561.793]	42.187 [0, 169.416]	147.988 [34.085, 412.491]	222.108 [53.133, 653.268]	9300.72 [3357.665, 24438.5]
0.6	1433.306 [177.734, 4216.765]	178.206 [48.291, 423.994]	498.371 [420.913, 683.139]	53.127 [0, 194.524]	133.185 [29.324, 391.78]	201.593 [48.488, 614.331]	8721.828 [3286.094, 23429.2]
0.7	1319.165 [160.266, 3969.552]	151.053 [28.35, 400.333]	615.818 [537.991, 803.129]	61.213 [0, 212.597]	121.045 [25.379, 354.014]	184.813 [46.268, 583.977]	8238.89 [3296.85, 21886.04]
0.8	1241.221 [140.577, 3852.836]	130.869 [15.312, 373.279]	737.362 [655.095, 945.976]	68.981 [0, 229.575]	113.259 [23.918, 346.094]	174.411 [45.045, 563.976]	7984.03 [3223.793, 21380.3]
0.9	1186.469 [140.532, 3710.653]	116.138 [8.078, 343.806]	861.692 [775.449, 1087.007]	75.392 [0, 227.153]	107.553 [20.807, 330.13]	167.62 [42.153, 518.69]	7839.05 [3242.567, 19986.49]

Table D9

Mean Values and 95% Confidence Intervals for the Cost Components Under Policy 2 for Varying Levels of the Parallelizable Portion of Surgeries for $c_W = 1$; $c_I = 50$; $c_O = 75$

q	Total Waiting Time Before PreIncision	Total Waiting Time Before Incision	Total Surgeon Idle Time	Total OR Idle Time	Total Surgeon Overtime	Total OR Overtime	Total Weighted Cost
0.1	3601.92 [1581.437, 6947.119]	765.641 [565.257, 1077.315]	195.565 [63.76, 362.917]	43.565 [0, 97.445]	470.815 [244.013, 805.047]	756.481 [398.568, 1267.93]	59499.46 [32625.58, 98471.33]
0.2	3562.289 [1743.99, 6494.863]	625.216 [450.213, 911.637]	232.997 [124.267, 399.491]	11.029 [0, 52.818]	375.739 [167.661, 693.031]	583.836 [260.449, 1069.055]	48526.66 [22364.2, 88052.35]
0.3	2780.171 [1035.543, 5639.578]	468.318 [323.321, 732.568]	262.049 [176.665, 425.593]	12.92 [0, 57.273]	275.82 [103.448, 574.543]	431.728 [147.698, 890.714]	36274.09 [13882.09, 73643.5]
0.4	2156.271 [546.529, 5010.821]	338.355 [183.604, 603.796]	314.69 [239.624, 489.512]	22.559 [0, 104.847]	204.522 [59.87, 495.546]	318.344 [74.565, 780.134]	27498.39 [9489.334, 64296.77]
0.5	1821.918 [361.888, 4619.97]	260.411 [106.474, 516.81]	404.675 [328.071, 591.098]	35.877 [0, 151.878]	167.923 [42.718, 443.555]	258.775 [58.443, 697.7]	23284.28 [8185.638, 58049.86]
0.6	1638.157 [285.375, 4295.699]	212.844 [71.705, 455.849]	512.651 [435.096, 699.343]	45.989 [0, 179.717]	147.59 [35.1, 393.989]	225.224 [53.384, 631.989]	21042.23 [7865.852, 52632.3]
0.7	1521.923 [245.331, 4280.565]	180.434 [46.803, 441.541]	629.078 [548.746, 836.187]	54.604 [0, 199.602]	134.78 [29.865, 391.471]	204.653 [48.068, 627.762]	19781.55 [7532.8, 52511.88]
0.8	1445.678 [229.303, 4159.767]	156.183 [29.344, 408.347]	750.436 [665.052, 968.566]	63.363 [0, 214.859]	126.745 [28.417, 368.633]	192.248 [46.724, 594.556]	19188.63 [7578.582, 49720.37]
0.9	1372.981 [201.568, 4024.81]	135.655 [15.533, 363.986]	871.789 [778.851, 1094.13]	69.482 [0, 221.191]	118.424 [26.25, 345.691]	180.239 [43.885, 552.603]	18500.69 [7335.695, 46529.56]