12-2022

Essays On Perioperative Services Problems In Healthcare

Amogh S. Bhosekar
Clemson University, abhosek@clemson.edu

ESSAYS ON PERIOPERATIVE SERVICES PROBLEMS IN HEALTHCARE

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Ennginееring

by
Amogh Bhosekar
December 2022

Accepted by:
Dr. Tugce Isik, Committee Chair
Dr. Sandra Eksioglu
Dr. Kevin Taaffe
Dr. Amin Khademi

Acknowledgments

I am incredibly thankful to my committee chair Dr. Tugce Isik and co-advisor Dr. Sandra Eksioglu for the invaluable guidance, continuous support, and tutelage during the course of my Ph.D. study. With their immense knowledge and expertise in various fields they have provided me with insightful comments and suggestions at every stage of the research that helped me improve my research work, writing, and presentation skills. They provided me with luxury to implement my novel thoughts that sometimes led to setbacks and sometimes to successful contributions.

Next, I would like to thank the dissertation committee members, Dr. Amin Khademi and Dr. Kevin Taaffe who provided valuable insights and guidance. They were immensely helpful and without their contributions and feedbacks during comprehensive exam and proposal defense, this dissertation would not be possible. They provided me with constructive criticism that was beneficial in improving my overall research. I would also like to thank everyone in the Industrial Engineering department including all professors and staff members for supporting me on this journey and creating a wonderful environment in the office.

I would also like to thank Prisma Health, located in Greenville, South Carolina, for funding part of my research and providing me with an opportunity to solve real-world business problems. I want to especially thank Dr. Robert Allen of Clemson University for his support and for helping me with data and providing access to conduct pilot studies in the hospital during the difficult time of COVID pandemic.

I want to acknowledge Clemson University's generous allotment of computing time on the Palmetto Cluster, which helped me run the experimental studies smoothly. I would like to thank Mr. Grigori and Mr. Asher for helping me when required with Palemetto. I would like to thank my manager at Ford Motor Company, Jun Park, for supporting me throughout and allowing me to take time off when needed.

Finally, I would like to thank my family in india and my friends who have stayed by my side during my ups and downs and have been constant pillars of support.

Table of Contents

List of Tables

List of Figures

Preface

One of the critical challenges in healthcare operations management is to efficiently utilize the expensive resources needed while maintaining the quality of care provided. Simulation and optimization methods can be effectively used to provide better healthcare services. This can be achieved by developing models to minimize patient waiting times, minimize healthcare supply chain and logistics costs, and maximize access. In this proposal, we study some of the important problems in healthcare operations management. More specifically, we focus on perioperative services and study scheduling of operating rooms (ORs) and management of necessary resources such as staff, equipment, and surgical instruments. We develop optimization and simulation methods to coordinate material handling decisions, inventory management, and OR scheduling.

In Chapter 1 of this dissertation, we investigate material handling services to improve the flow of surgical materials in hospitals. The ORs require timely supply of surgical materials such as surgical instruments, linen, and other additional equipment required to perform the surgeries. The availability of surgical instruments at the right location is crucial to both patient safety and cost reduction in hospitals. Similarly, soiled material must also be disposed of appropriately and quickly. Hospitals use automated material handling systems to perform these daily tasks, minimize workforce requirements, reduce risk of contamination, and reduce workplace injuries. Most of the literature related to AGV systems focuses on improving their performance in manufacturing settings. In the last 20 years, several articles have addressed issues relevant to healthcare systems. This literature mainly focuses on improving the design and management of AGV systems to handle the specific challenges faced in hospitals, such as interactions with patients, staff, and elevators; adhering to safety standards and hygiene, etc. In Chapter 1, we focus on optimizing the delivery of surgical instrument case carts from material departments to ORs through automated guided vehicles (AGV). We propose a framework that integrates data analysis with system simulation and optimization. We test the performance of the proposed framework through a case study developed using data from a partnering hospital, Greenville Memorial Hospital (GMH) in South Carolina. Through an extensive set of simulation experiments, we investigate whether performance measures, such as travel time and task completion time, improve after a redesign of AGV pathways.

We also study the impact of fleet size on these performance measures and use simulation-optimization to evaluate the performance of the system for different fleet sizes. A pilot study was conducted at GMH to validate the results of our analysis. We further evaluated different policies for scheduling the material handling activities to assess their impact on delays and the level of inventory required. Reducing the inventory level of an instrument may negatively impact the flexibility in scheduling surgeries, cause delays, and therefore, reduce the service level provided. On the other hand, increasing inventory levels may not necessarily eliminate the delays since some delays occur because of inefficiencies in the material handling processes. Hospitals tend to maintain large inventories to ensure that the required instruments are available for scheduled surgery. Typically, the inventory level of surgical instruments is determined by the total number of surgeries scheduled in a day, the daily schedule of surgeries that use the same instrument, the processing capacity of the central sterile storage division (CSSD), and the schedule of material handling activities. Using simulation-optimization tools, we demonstrate that integrating decisions of material handling activities with inventory management has the potential to reduce the cost of the system.

In Chapter 2 we focus on coordinating OR scheduling decisions with efficient management of surgical instruments. Hospitals pay more attention to OR scheduling. This is because a large portion of hospitals' income is due to surgical procedures. Inventory management of decisions follows the OR schedules. Previous work points to the cost savings and benefits of optimizing the OR scheduling process. However, based on our review of the literature, only a few articles discuss the inclusion of instrument inventory-related decisions in OR schedules. Surgical instruments are classified as (1) owned by the hospital and (2) borrowed from other hospitals or vendors. Borrowed instruments incur rental costs that can be up to 12-25% of the listed price of the surgical instrument. A daily schedule of ORs determines how many rental instruments would be required to perform all surgeries in a timely manner. A simple strategy used in most hospitals is to first schedule the ORs, followed by determining the instrument assignments. However, such a strategy may result in low utilization of surgical instruments owned by hospitals. Furthermore, creating an OR schedule that efficiently uses available surgical instruments is a challenging problem. The problem becomes even more challenging in the presence of material handling delays, stochastic demand, and uncertain surgery duration. In this study, we propose an alternative scheduling strategy in which the OR scheduling and inventory management decisions are coordinated. More specifically, we propose a mixed-integer programming model that integrates instrument assignment decisions with OR scheduling to minimize costs. This model determines how many ORs to open, determines the schedule of ORs, and also identifies the instrument assignments for each surgery. If the level of instrument inventory cannot meet the surgical requirements, our model allows instruments to be rented at a higher cost. We introduce and evaluate the solution methods for this problem. We propose a Lagrangean decomposition-based heuristic, which is an iterative procedure. This heuristic separates the scheduling problem from the inventory assignment problem. These subproblems are computationally easier to solve and provide a lower bound on the optimal cost of the integrated OR scheduling problem. The solution of the scheduling subproblem is used to generate feasible solutions in every iteration. We propose two alternatives to find feasible solutions to our problem. These alternatives provide an upper bound on the cost of the integrated scheduling problem. We conducted a thorough sensitivity analysis to evaluate the impact of different parameters, such as the length of the scheduling horizon, the number of ORs that can be used in parallel, the number of surgeries, and various cost parameters on the running time and quality of the solution. Using a case study developed at GMH, we demonstrate that integrating OR scheduling decisions with inventory management has the potential to reduce the cost of the system.

The objective of Chapter 3 is to develop quick and efficient algorithms to solve the integrated OR scheduling and inventory management problem, and generate optimal/near-optimal solutions that increase the efficiency of GMH operations. In Chapter 2, we introduced the integrated OR scheduling problem which is a combinatorial optimization problem. As such, the problem is challenging to solve. We faced these challenges when trying to solve the problem directly using the Gurobi solver. The solutions obtained via construction heuristics were much farther from optimality while the Lagrangean decomposition-based heuristics take several hours to find good solutions for large-sized problems. In addition, those methods are iterative procedures and computationally expensive. These challenges have motivated the development of metaheuristics to solve OR scheduling problems, which have been shown to be very effective in solving other combinatorial problems in general and scheduling problems in particular. In Chapter 3, we adopt a metaheuristic, Tabu search, which is a versatile heuristic that is used to solve many different types of scheduling problems. We propose an improved construction heuristic to generate an initial solution. This heuristic identifies the number if ORs to be used and then the assignment of surgeries to ORs. In the second step, this heuristic identifies instrument-surgery assignments based on a first-come, first-serve basis. The proposed Tabu search method improves upon this initial solution. To explore different areas of the feasible region, we propose three neighborhoods that are searched one after the other. For each neighborhood, we create a preferred attribute candidate list which contains solutions that have attributes of good solutions. The solutions on this list are evaluated first before examining other solutions in the neighborhood. The solutions obtained with Tabu search are compared with the lower and upper bounds obtained in Chapter 2. Using a case study developed at GMH, we demonstrate that high-quality solutions can be obtained by using very little computational time.

Chapter 1

Simulation Optimization of Material Handling Activities and Inventory Management in Hospitals

1.1 Introduction

Achieving better patient care is a key objective of any healthcare system. Although improving supply chain activities in hospitals is important, little research has been done in this area [41]. Patient needs should not be neglected when patient care is coordinated with supply chain activities to improve efficiency. Material handling activities such as delivering food, medication, and clean linen to patients and removing waste in a timely manner require resources and coordination [106, 34]. Most of these activities are repetitive, occur several times a day, and use a large portion of the labor hours of a healthcare provider. About 45% of the respondents to a Cardinal Health survey conducted in 2019 indicated that manual supply chain tasks have a negative impact on patient care. Approximately 42% of the respondents to this survey acknowledged that tasks related to the supply chain take a lot of time away from patient care [1]. Furthermore, healthcare providers also spent twice the amount of time required on supply chain-related tasks. Inefficiencies in supply chain operations contribute to higher healthcare care costs in industrialized nations. Supply chain costs are the second largest expense in hospitals, after personnel costs, and represent about 15 to 30% of total net patient revenues [159]. The researchers estimate that logistics-related expenses account for up to 40% of operating budgets in hospitals [40]. According to a Tecsys white paper, inbound logistics account for only 20% of the total supply chain activities in hospitals and clinics. The remaining 80% activities focus on internal logistics that support services such as food trays, IT equipment, laboratory supplies, records, linen and laundry, etc. About 95% of a healthcare system's supply chain costs of a healthcare system is due to the management of internal logistics [144]. Automation in healthcare is used to improve the efficiency of material handling systems and reduce these operational costs. Many hospitals use automated guided vehicles (AGVs) for repetitive tasks, as the use of AGVs leads to reduced material handling costs and reduced liabilities due to workplace injuries (from lifting and moving heavy weights) [120]. The use of AGVs also reduces the risk of contamination of sterile instruments, food, etc. The work by [85] provides detailed guidelines for the design and implementation of automated systems in healthcare and a discussion of prototypes and/or products.

The first part of this chapter addresses the challenges associated with AGV delivery systems at our partner hospital, GMH. More specifically, this part focuses on the delivery of surgical case carts to GMH ORs using automated guided vehicles (AGV). At GMH, an AGV system has already been installed and several rules are outlined for pedestrian traffic safety. For years, the material handling process at GMH has not changed. However, to maintain the level of service provided to an increasing number of patients, additional AGVs were added to the system without updating the physical infrastructure. This resulted in an increase in the congestion of the AGVs. Furthermore, AGVs are not allowed to pass each other; thus, if for some reason an AGV stops, the other AGVs that follow it also stop at a safe distance, contributing to traffic. The congestion of AGVs leads to material handling delays in GMH. To reduce these material handling delays, we propose a framework that integrates data analysis with system simulation and optimization.

Although one of our objectives is to improve material handling at GMH, the broader objective of this research is to develop solutions that increase the efficiency of operations at GMH by improving decisions regarding inventory, scheduling, and transportation of instruments. To do this, we conducted an extensive analysis of material handling processes. Based on our data analysis, it was worth evaluating the impact of changes in current AGV routes and the location of some GMH departments on congestion. This motivates our first research question (i) Can performance measures, such as travel time and task completion time for AGVs, be improved after a redesign of AGV pathways at GMH? To address this research question, we developed two simulation models that capture the movement of AGV before and after the redesign of the pathways. The two resulting AGV pathway designs were then compared using an extensive sensitivity analysis.

The problem of sizing the AGV fleet has been addressed in the literature in the installation stage of an AGV system. As the number of trips (volume of surgical cases) changes over time, there is increased congestion on existing pathways. Therefore, it is important to re-evaluate the ideal fleet size based on the volume of cases. This motivates our second research question (ii). Do performance measures, such as travel time and task completion time, improve when the number of AGVs used daily is controlled by the volume of surgical cases? If so, how many AGVs should be used daily? We used a simulation-optimization model to evaluate the performance of the system for different AGV fleet sizes. Finally, we conducted a pilot study at GMH to validate the results of our analysis. The result of this study indicates that the proposed solution, which uses a smaller AGV fleet than currently used in GMH, leads to significant reductions in congestion and travel times and increased AGV utilization.

Perioperative services department (PSD) oversees the timely delivery of surgical instruments to OR, transportation of soiled instruments back from OR, sterilization activities, and ensures that instruments are available for all surgeries [109]. Another important task of perioperative services is to efficiently manage the inventory of surgical instruments. Delays in the availability of instruments is one of the recurring sources of surgical delays [32]. Mitigating these delays requires close collaboration between the central sterilization and the OR staff [80]. A study by [162] suggests that 45.9% of the delays in an OR occur because an instrument was unavailable. These delays resulted in longer working hours for physicians and staff and, thus, additional costs for the hospital. The delay in surgery due to the lack of instruments also negatively affects quality of care and adverse effects can occur [162]. Some delays still occur due to inefficiencies in the material handling process. For example, congestion, due to the movement of Automated Guided Vehicles (AGVs) along the narrow corridors of a hospital, can cause delays. To our knowledge, few studies evaluate the effects of inventory and material handling decisions on the level of service provided by ORs in hospitals. This gap in the literature is the main motivation for this research.

The second part of this chapter addresses challenges in coordinating material handling with inventory management of surgical instruments. Material handling and inventory management decisions are critical to hospital and OR service levels and costs. However, efficiently coordinating these decisions is challenging due to their interdependence and the uncertainties faced by hospitals.

Surgeries performed in most hospitals are classified as elective or emergency [72]. The exact timing of elective surgery and its OR assignment is finalized between 24 and 48 hours before the day of surgery. The OR scheduler makes these assignments after considering the availability and preferences of the surgeon and surgical staff, as well as the availability of the required equipment and instruments. These assignments affect the availability of the instruments for the remainder of the day [72]. In GMH, an instrument is not used more than once on the same day. This is because the delivery of surgical instruments is only performed once a day. Surgical instruments are classified as: (1) owned by the hospital, (2) borrowed/rented from other hospitals or doctors, or (3) consigned by a vendor who owns the instrument [27]. Typically, a hospital would not purchase an instrument if it was only used in rare or specialty surgeries [27]. In such a case, hospitals consign or rent the instrument and pay the owner upon its use. A hospital has several other reasons to borrow instruments, such as to accommodate physicians' requests to schedule consecutive surgeries during a given day, continue operations on a limited budget, or mitigate a lack of storage space [134]. According to [134], such practices lead to inefficiencies because borrowed instruments create an additional workload for the sterile processing department (SPD) due to the requirements to maintain documentation and pack and sterilize instruments. In addition, some instruments have special cleaning procedures, which may differ from other procedures for hospital-owned instruments. Following these procedures increases the workload of the employees. Additionally, consigned instruments stored in the hospital occupy additional storage space. A recent study conducted in a major academic hospital in the US suggests that half of the instruments are consigned and their cost is on average 12% more than the instruments owned by the hospital [99]. These challenges motivate our third research question (iii) How does the inventory level of surgical instruments impact the service level provided by the ORs? We present a numerical study that evaluates how inventory levels impact instrument utilization and surgery delay.

Inefficiencies in material handling activities lead to delays that affect instrument availability. Furthermore, the duration of a surgery is uncertain; thus, a surgery may take longer than planned, making instruments unavailable. One of the reasons hospitals do not plan to reuse instruments is to avoid delays due to the unavailability of surgical instruments. For example, GMH plans to use an instrument in no more than one surgery a day. Instruments are delivered to a storage area beside the OR the night before surgery. This practice ensures that the necessary instruments are available during surgery. As a result, the same instrument cannot be reused in other surgeries scheduled on the same day. Alternatively, an instrument could be delivered directly from the CSSD a short time before the start of surgery, using the JIT delivery approach. This practice increases the utilization of instruments used in short-duration surgeries performed earlier in the day. This approach can also lead to lower inventory levels and lower inventory holding costs. However, such an approach requires coordination of material handling, instrument decontamination and sterilization, and OR scheduling. These challenges motivate our fourth research question (iv) How do material handling activities impact the level of service provided by ORs? A numerical analysis is conducted using the proposed simulation models to answer this question. We use the previously developed simulation models to develop three additional discrete event simulation models to address this question. Model 1, *Current*, assumes that there is no coordination of material handling and inventory decisions. Model 2, *Two Batch*, assumes partial coordination and Model 3, *Just-In-Time* (JIT) assumes full coordination. Each material handling approach follows a different schedule of case cart delivery to ORs. For each approach, a numerical study is conducted to assess how the number of AGVs affects travel time, congestion, utilization of AGVs, delivery time and instrument utilization.

Finally, the decision to reduce the inventory of an instrument limits the time that that instrument is available. This, in turn, negatively affects flexibility in scheduling a surgery that requires the instrument and therefore the level of service provided. The problem becomes even more challenging when coupled with an inefficient material handling system and uncertain surgery durations. These challenges motivate our fifth research question (v) How do integrating inventory management and material handling decisions impact the service level provided by ORs?

The following is an outline of this chapter. Section 2.2 reviews the literature relevant to this work and Section 2.3 provides a detailed description of the material handling system. Section 2.4 describes the proposed simulation model of the material handling system and introduces the case study to evaluate the impact of fleet sizing experiments. Section 2.5 provides a detailed description of the integrated material handling and inventory management problem. This section introduces additional simulation models and another case study that evaluates the effect of coordination of material handling activities with inventory management. Finally, Section 2.6 presents the final remarks.

1.2 Literature Review

The main stream of existing literature relevant to this work is inventory management of reusable surgical instruments. Since AGVs are used as carriers by this study's partner hospital, as well as in many others, the literature that discusses the use of AGV systems for material handling in hospitals is also reviewed. Other streams of research which are related to this work are optimization through simulation [58, 122, 8, 139] and JIT production systems [67, 89, 76, 123, 124].

Automated material handling systems have been widely studied in manufacturing settings. The existing literature related to AGV systems focuses mainly on selection of fleet size and design of AGV systems. Work by [104], [46], [143], [137], [108], [126], [10], [29], [18] use simulation and optimization models to identify the size of the AGV fleet. Work by [59, 73] focuses on operational issues that arise in repetitive manufacturing systems with unidirectional material flow. [59] investigates flow-path design, fleet sizing, job and vehicle scheduling, dispatching, and conflict-free routing. [73] proposes AGV dispatching policies to maximize the throughput rate. [92] provide a review of the literature related to the design and control of AGV systems. This review identifies the following key factors that impact the design of these systems: the required number of AGVs, the AGV schedule, the AGV route, the idle position, etc. The work by [154] provides another review of the literature related to the design and control of AGV systems used in manufacturing, distribution, and transportation.

Four types of approaches are adapted to determine the size of the AGV fleet for a logistics network: (a) calculus-based approaches; (b) deterministic optimization approaches; (c) stochastic optimization approaches; and (d) simulation-based approaches [28]. Early work in calculus-based models focuses on empty and loaded travel times of AGVs [45, 55]. However, travel time depends on congestion, which is affected by facility layout, vehicle speed, and load size [153]. The work of [97] uses analytical models to assess the impact that increasing the flexibility of the AGV routes has on the number of AGVs needed. However, analytical models fail to capture congestion in the system [59].

The work of [9] develops a regression model to determine the size of the fleet as a function of the number of work centers, the lengths of the routes, and the number of intersections. In contrast to simulation models, these regression models provide quick results, but do not capture the complexities of the transportation system, such as the movement of elevators and the movement of AGVs along shared paths.The use of optimization methods to improve material handling systems is common in the literature. Deterministic methods, such as integer programming, multi-objective optimization, and mixed integer programming, are used to model system dynamics [136, 103, 125]. For example, a minimum cost flow model is developed to determine the minimum number of vehicles required in a container terminal [155]. Work by [83] develops an analytical model to estimate an upper and a lower bound on the number of vehicles required in a transportation system. Work by [125] proposes analytical and simulation methods to determine the number of vehicles required. Their analytical method is based on load handling time, empty travel time, waiting time, and blocking time. The authors also proposed a mixed-integer program with the objective of minimizing empty trips subject to a limited number of trips to and from each load transfer station. Simulation methods are used to validate initial estimates of the size of the fleet. The work of [45] discusses the performance of non-simulation-based approaches to determine the size of the fleet under different dispatch rules. These approaches are observed to underestimate the required fleet size compared to simulation-based approaches. Some researchers utilize stochastic approaches, such as using queueing models to minimize the number of AGVs used. The steady-state behavior of closed queueing networks can be used to estimate the required fleet size. The results of these methods can be compared with simulation models for validation purposes [142, 28]. A hierarchical queueing approach has also been used to determine how many vehicles are required [100].

Simulation-based approaches are considered time-consuming and costly [45, 153]; however, they can handle the complexities and randomness present in real systems. Therefore, simulation optimization models have previously been used to model complex inventory replenishment problems [82], medical supply chain problems [115], and fleet size problems to understand the performance of the system. According to [45], the analytical methods he developed under estimated the requirement (vehicle) in most dispatch strategies. His methods gave close estimates in only one of the strategies. According to [142], the analytical methods help determine the starting point for the number of vehicles to be used in a simulation experiment. A review of the literature by [59] indicates that simulation studies are promising tools to estimate the size of the fleet. Often, simulation models use analytical tools to provide estimates of fleet size, which serve as a starting point for the simulation. The work of [129] uses simulation to model a complex AGV system. This model provides the flexibility needed to analyze the systems due to the complexities present in modeling the AGV system. Although analytical methods could be used to determine the number of AGVs in some systems, they are not a good fit due to shared paths and they would not provide performance measures related to congestion that can be compared to current practice and the alternative designs considered for the material handling system. Models based on idle and waiting times for machines, parts, and AGVs are also being developed, as well as number and speed of vehicles [90, 93]. Other studies use two-stage approaches for system simulation and evaluation, or develop case studies to determine the size of the fleet and evaluate the impacts on performance indicators, such as queue sizes, occupation numbers, and service times [60, 165].

The literature focused on the design of AGV systems for healthcare applications has grown over the past 20 years. [85] discusses several factors that must be considered to design a mobile robotic system for healthcare applications, and his work provides several guidelines for researchers to improve these designs. Simulation models are developed to evaluate the performance of AGV-used material handling systems. These models are often used to compare automated systems with manual delivery systems [130, 129, 25]. For example, a study by [129] compares the performance of a manual and an automated material handling system that uses AGVs for the delivery of clinical supplies and pharmaceuticals in a hospital. Costs, turnaround time, variability of turnaround time, cycle time, and utilization are used as performance measures. The proposed simulation model shows that the use of robotic delivery is economically viable and improves the performance measures listed above. Another study, [130] uses the analytical hierarchy process to build a decision problem that evaluates the performance of a robotic healthcare delivery system based on technical, economic, and several other factors. Their proposed simulation model assesses the technical factors that include the speed of robots and human couriers based on the arrival rates of visitors who request the elevator, the availability of the elevator, the arrival rates of the delivery items that request robots, and the availability of robots. The work of [25] compares three supply chain models that use: a) manual inventory check and delivery, b) RFID inventory check and manual delivery, and c) manual inventory check with AGV-based material handling. This study shows that the use of AGVs for material handling is economically viable, maximizes cost savings, and produces ergonomic benefits due to reduced manpower requirements. Through a simulation-based case study, [120] identify the potential benefits of using an AGV system in a hospital. The work of [57] proposes a data-driven agent-based simulation model to analyze the current status of the goods delivery system and identify potential countermeasures to improve internal logistics. The work of [17] addresses the gap that exists in the literature between the technological aspects of automation, organizational issues related to automation, and management of hospital logistic staff. Their results highlight the need for new knowledge and skills to improve the design and management of AGV systems in hospitals.

The cost of ORs is affected by the availability of surgical supplies and implants [23]. Surgical supplies include soft goods and instruments required for surgery. These supplies can be reused or disposable. Numerous studies show that the cost of reusable supplies is significantly lower than the cost of disposable supplies [37, 44, 133, 98, 2], and disposable supplies negatively impact the environment [2]. [23] conduct a study to assess the cost of opened and unused soft goods and instruments in a French hospital. They reported that wasted supplies have a median cost of €4.1 per procedure, which represents approximately 20.1% of the cost of surgical supplies. However, most hospitals do not have standardized procedures to manage surgical supplies inventory [3]. In their review paper, [3] focuses on the inventory management of sterile instruments. Notice that, most surgical instruments used in surgery are grouped into containers called instrument trays. This work identifies three important considerations for inventory management: assignment of instruments and quantity for each type of tray, assignment of the type of tray to a surgeon or procedure, and the number of trays carried by the hospital. Decisions related to the first two considerations are affected by the surgeon's preferences, indicated in the DPC.

The cost of surgical supplies can be reduced in several ways, including 1) improving the accuracy of DPCs, 2) increasing surgeon awareness, and 3) standardizing surgical techniques. The accuracy of the DPC can be improved by periodically reviewing it [74, 51] or by recording which instruments are used on a tray and removing the instruments that are not used [113, 43]. For example, [74] show that the participation of physicians in the review of the corresponding DPC led to the removal of 109 disposable supplies and the elimination of 3 reusable instrument trays. Consequently, the cost of a case cart was reduced by $16 on average. According to a survey conducted by [81], surgeons often underestimate the cost of expensive items and overestimate the cost of less expensive items due to internal bias and ignorance of costs [81]. Therefore, the cost of a surgical procedure can be reduced by increasing the awareness of surgeons of standardized operating equipment and the cost of instruments [62, 14]. Finally, the work by [138] shows that standardization of surgical techniques can significantly reduce operating costs without affecting the quality of a procedure [138]. [140] indicate that the tailored and streamlined tray compositions lead to significant cost savings [[140]]. Furthermore, surgeons prefer trays with fewer unsolicited instruments [41, 140]. Several optimization models have been developed to solve the tray optimization problem and address tray composition and inventory management for reusable surgical instruments. The objective of this problem is to minimize an OR's cost by optimizing the number of trays utilized and the amount of inventory supplied. The problem also addresses the preferences of surgeons for instruments. [41] develop a linear integer programming formulation and propose a heuristic algorithm to obtain a solution to this problem [41]. [127] propose a resource sharing method for reusable devices. The objective is to minimize the storage, processing, and waste costs of supplies that have not been used. [147] propose a deterministic model that minimizes the cost of instrument storage and delivery by optimizing the composition of the tray. [5] present a bi-objective optimization model for the configuration of surgical trays with ergonomic considerations. The first objective function minimizes the total number of types of assembled tray, and the second objective function minimizes the total number of instruments that were not requested. They used the ε- constraint method to obtain the Pareto-optimal front. [42] develop an exact integer linear programming formulation, a row and column generation approach, a greedy heuristic, and several meta heuristics to solve tray optimization problem. These approaches are evaluated on the basis of the average computation time, the average value of the objective function, and the number of solutions for which optimality is proven.

Our proposed research framework enables hospitals to identify the factors that affect the performance of the material handling system and develop solutions that improve its efficiency. Previous work points to the cost savings and benefits of using simulation to model AGV movements. However, based on our review of the literature, only a few articles discuss the use of AGVs in hospitals [119, 25, 56, 129]. The research papers cited here treat the vehicle fleet sizing problem as a tactical issue to be addressed at the design stage, but the problem becomes operational when one selects the required number of vehicles from a pool of vehicles on a day-to-day basis. Our work addresses operational-level issues associated with fleet size selection that impact AGV movements. In addition, most of the literature related to AGV systems focuses on improving their performance in manufacturing settings. In the last 20 years, several articles have addressed issues relevant to healthcare systems. This literature mainly focuses on improving the design and management of AGV systems to handle the specific challenges faced in hospitals, such as interactions with patients, staff and elevators; adhering to safety standards and hygiene, etc. Our work highlights the role of coordinating decisions between material handling and inventory management in improving the level of service provided by ORs. In particular, this work demonstrates that JIT delivery of surgical cases for short-duration surgeries can potentially improve the level of service provided by ORs, thus reducing the cost of healthcare. We also developed a real-life case study using data from a US-based hospital. The proposed material handling approaches, which are intuitive and easy to implement, are verified and validated using historical data. The results of the proposed analysis have inspired the partner hospital featured in this study to improve material handling and inventory management practices. Although other healthcare facilities may not choose to implement the models presented here, they can learn from these practices.

1.3 Description of the System

The research presented in this chapter was conducted in collaboration with GMH, one of the seven campuses of Prisma Health in South Carolina, USA. GMH provides general inpatient services and specialized treatments for heart disease and cancer. The hospital also houses the Family Birthplace, the Children's Hospital, and the Children's Emergency Center.

This research focuses on material handling activities that support surgical processes at GMH. The research team collaborated with the Perioperative Services Department (PSD) which oversees these processes. PSD consists of three divisions: Materials Division (MD) and Central Sterile Storage Division (CSSD), both of which are located on the mezzanine floor (see Figure 1.1), and Operating Room Division (ORD) located on the second floor. The second floor contains 32 operating rooms (ORs) divided into three separate cores. The cores are grouped according to the medical specialties they serve, such as orthopedic, cardiovascular, and neurological treatment. The instruments used in a surgery are stored in the corresponding core.

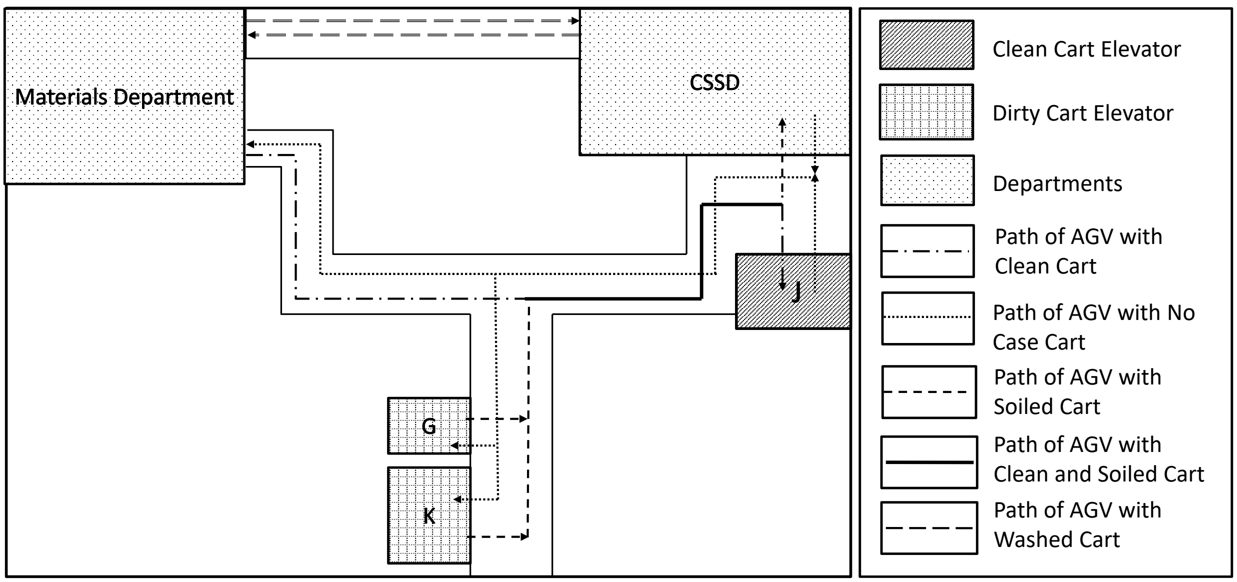


Figure 1.1: Map - Mezzanine Floor.

The type of surgery determines the materials needed, including soft goods and implants, and the surgical instruments used. The PSD is responsible for loading materials into a clean case cart; delivering the case cart from the MD to the OR; loading implants and instruments to clean case carts in the cores; delivering clean case carts from the cores to the OR; returning soiled instruments, which have already been used, from the OR to the CSSD; and cleaning the soiled instruments in the CSSD. Each case cart is dedicated to a particular surgical case and contains all the material requested by the surgeon. AGVs manufactured by FMC-Technology are used in GMH to move clean and soiled case carts. Figure 1.1 shows the AGV routes and the location of the departments.

The PSD-managed material handling process starts with the OR manager providing a detailed schedule of the surgeries planned for the next day. Based on the OR schedule and the doctors' preferences, a list of instruments and soft goods is generated in the MD. From 3 pm, soft goods are manually loaded onto clean case carts. This stage is called *picking process*. The carts are then manually moved to the detents. Detents are platforms or areas equipped with the rails necessary to load and unload an AGV. Next, a request for an AGV is submitted through a centralized AGV control system and an available AGV, closest to the MD, is assigned to the case cart. The movement of a loaded AGV is depicted in Figure 1.1 as *"Path of AGV with Clean Cart.?* This AGV uses the elevator J to move the cart to the second floor. The clean case cart is then dropped off at one of the detents in the case cart storage area (CCSA) located next to elevator J on the second floor.

Once every case cart has been delivered to the CCSA, an inspection is performed to ensure that the required soft goods are delivered. If clean case carts are not delivered by 7 pm, the hospital incurs overtime. Additionally, GMH uses AGVs to transport dietary and linen carts, both of which have higher priority than surgical case carts. The movement of these carts begins at 6:00 pm and their deliveries will be completed on the same night. As a result, AGVs become increasingly unavailable for the movement of surgical case carts after 6 pm. Therefore, the delivery of surgical case carts must be completed before other services begin to request AGVs for transport.

The case carts are stored in the CCSA until the next day, the day of surgery. Instruments and implants, which are stored in one of the cores, are added to the case cart. The case cart is then manually moved to the OR. After surgery, the cart is considered soiled and must be decontaminated. A soiled cart is manually moved to the detents on the second floor and an AGV is requested on the centralized AGV control system. The assigned AGV moves the dirty cart to the CSSD. The movement of this AGV is depicted in Figure 1.1 as *"Path of AGV with Soiled Cart"* Then, the AGV uses the elevator G or K to move the cart to the mezzanine floor. The portion of the path that is *shared* by AGVs with clean and soiled case carts is shown in Figure 1.1 as *"Path of AGV with Clean and Soiled Cart"*.

Soiled instruments are washed and sterilized at CSSD to comply with safety guidelines, and sterilized instruments are loaded into a clean case cart and moved to the corresponding core for storage. The soiled case carts are washed in the cart washer. Once the cart is clean, an automatic request for an AGV is sent to pick up the washed cart. The movement of AGVs with cleaned carts is shown in Figure 1.1 as "*Path of AGVs with Washed Cart"*. The washed carts are dropped off at the MD for the picking process. This cycle of surgical case carts starts and ends in the MD and is repeated every business day.

For years, material handling for perioperative service processes at GMH has not changed. However, in recent years, the number of patients served by GMH has increased rapidly. In an effort to improve the services provided, additional AGVs and case carts were added to the system without updating the physical infrastructure. As a result, GMH's staff noticed that AGVs loaded with case carts often sit on the mezzanine floor waiting for elevator J. AGVs from the CSSD and elevator J have a higher priority than AGVs that move toward these locations. Therefore, AGVs traveling to these locations wait for elevator J for a long time. Furthermore, AGVs are not allowed to pass each other; therefore, if for some reason an AGV stops, the other AGVs that follow will also stop at a safe distance, contributing to traffic. Congestion leads to a shortage of AGVs in the MD. Sometimes, the soiled case carts are stuck in traffic, creating a shortage of washed case carts and clean instruments. These shortages lead to further delays in delivering clean carts.

Occasionally, the cart washer and instrument washer at the CSSD remain idle for longer periods of time, contributing to under-utilization of the equipment. In 2017, PSD staff reached out and expressed their interest in a study of the current material handling system. The GMH team wanted to know how a change in current AGV routes and the location of some GMH departments would impact congestion. Based on the layout of the mezzanine floor, it seems intuitive that changing the roles of elevators G and K with J would lead to less congestion. We conducted an extensive data analysis of material handling processes. Based on the results, we decided to also investigate the possible impacts of reducing the number of AGVs moving surgical case carts. To this end, two simulation models were developed.

Data collection and analysis:

Data were collected and analyzed to understand the system, discover inefficiencies, and support discrete event simulation models. We obtained data from the AGV control system for 50 consecutive days. Data provided information on the movement of AGVs, such as the date, time, and location of the pickup; date, time, and location of the drop-off; and type of cart an AGV is carrying. Since the scope of the study is limited to surgical services, only data on surgical case cart movement are analyzed. In the event of an AGV breakdown, the AGV is moved from its path and moved to a maintenance area. The data points for those AGVs are removed from the data set as outliers. Additionally, AGVs stop if there is a person or another obstacle, such as another AGV, in their scanning radius. Micro-stoppages due to human traffic are not recorded separately in the data and are negligible compared to travel times. For this reason, these micro-stoppages are not modeled explicitly. The details on how micro-stoppages due to AGVs traffic are modeled are explained in the next section as part of our modeling approach.

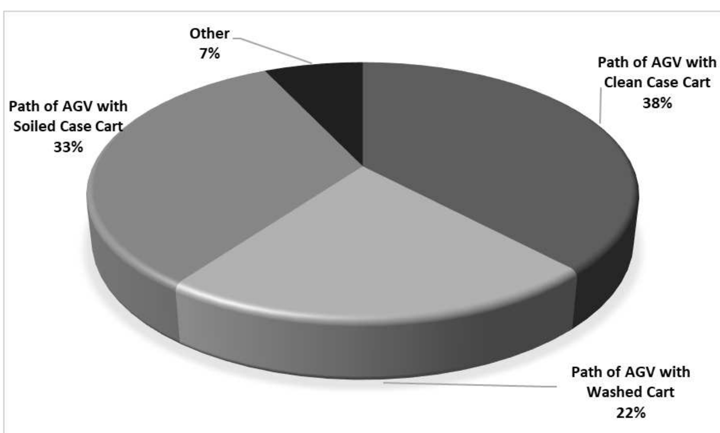


Figure 1.2: Number of Trips by Route

Figure 1.2 shows the number of AGV movements along each route. 93% of the movements are associated with the routes shown in Figure 1.1, that is, on the mezzanine floor. Further analysis indicates that 71% of these movements use the "*Path of AGV with Clean and Soiled Cart*". Since most of the movements occur on these routes, the data analysis focuses on only two routes that contribute the most to traffic congestion, "*Path of AGV with Clean Cart*" and "*Path of AGV with Soiled Cart*". For each cart, the travel time along both routes is calculated based on the data on drop-off and pick-up times. These travel times were grouped over 3 to 4 hour intervals, since this is how long the picking process takes on average. The average and standard deviation of the travel times along each path at different times of the day are summarized in Table 1.1. These results indicate that travel times are longest along the "*Path of AGV with Clean Cart*" during 3 pm to 7 pm, when the picking process takes place.

Table 1.1: AGV Movements by Time of Day

Route	Time Interval	No. of Trips	Travel Time [Min]		Coefficient of Variation
			Average	Std. Dev.	
2nd Floor Soiled Cart Storage - CSSD	12am-3am	44	7.28	6.16	0.85
	3am-6am	53	6.87	6.16	0.9
	6am-9am	146	5.55	3.27	0.59
	9am-12pm	981	4.56	4.04	0.88
	12pm-3pm	882	5.17	2.56	0.49
	3pm-7pm	753	**9.71**	**8.51**	0.88
	7pm-9pm	126	6.65	3.04	0.46
	9pm-12am	77	5.45	1.01	0.19
Materials Department - Case Cart Storage	12am-3am	131	4.66	10.26	2.2
	3am-6am	227	5.96	9.55	1.6
	6am-9am	112	5.27	6.17	1.17
	9am-12pm	80	5.8	2.4	0.41
	12pm-3pm	101	5.33	3.69	0.69
	3pm-7pm	**1416**	**8.94**	**6.49**	0.73
	7pm-9pm	254	5.67	4.45	0.78
	9pm-12am	196	4.88	5.72	1.17

The results of the data analysis generated the input parameters used in the simulation model. For example, each trip along "*Path of AGV with Clean Cart*" represents a surgery scheduled for the next day. The total number of surgical cases differs by day of the week, that is, Monday to Friday. The number of trips for each day is summarized for each week during this period. This gives seven data points for each day of the week. Due to the limited number of data points, the triangular distribution (TRIA) is used to represent the total number of surgeries scheduled per day. To derive this distribution, the minimum and maximum number of surgeries is determined and the mode for each day of the week is estimated. The corresponding results are summarized in Table 1.2. Data for AGV trips on the "*Path of AGV with Soiled Cart*" is used to estimate the release time of soiled carts from the operating room, as soiled carts are delivered to the CSSD immediately after surgery. Data on the total number of soiled carts delivered at the end of every half hour for each day of the week are used to distribute the total number of surgeries over different time intervals within a day. Other input parameters used include the number of AGVs used for the movements of surgical case carts, the number of case carts available, and the number of cart washers in the system.

Table 1.2: Total Number of Surgeries per Day

Day	Distribution
Mon.	TRIA (60,68,75)
Tue.	TRIA (65,72,76)
Wed.	TRIA (60,65,72)
Th.	TRIA (69,75,80)
Fri.	TRIA (55,62,69)

Note that the existing pathways for the movement of AGVs with clean and soiled carts are influenced by safety regulations and the movement of other carts. For example, elevators G and K are used to move soiled surgical instruments, dirty linen, and trash. These elevators continue to the basement to deliver dirty linen and trash. On the other hand, elevator J is only used for the movement of clean surgical instruments to eliminate any potential contamination. Thus, this elevator serves only the mezzanine and the second floor.

1.4 Simulation-optimization of Material Handling Activities at GMH

First, a conceptual model is developed based on the framework proposed by [128]. This conceptual model is then used in the development of the DES model for the rest of this chapter. The details of these models are presented below.

1.4.1 Conceptual Model

The framework proposed by [128] outlines the steps to develop a conceptual model. We followed these steps and determined: (*i*) organizational aims (see Figure 1.3); (*ii*) modeling objective (see Figure 1.3); (*iii*) project objectives (see Figure 1.3); (*iv*) scope of the model (see Table 3 in the Appendix); (*v*) level of detail in the model (see Tables 4 and 5 in the Appendix); (*vi*) model assumptions (see Table 1.4); and (*vii*) model simplifications (see Table 1.4).

Movement of AGVs and elevators We develop a guided path transporter network to model AGV movement on the mezzanine floor. This guided path functions as a physical entity in the simulation and is used to model the traffic flow. The data necessary to model AGV movements, such as velocity, acceleration, deceleration, and turning velocity, are obtained from the FMC-Technology AGV handbook. The data necessary to model elevators, such as the time it takes to open and close the door and dimensions, are obtained from the same handbook. The length of links in the transporter network is calculated using GMH floor maps.

The AGV network consists of intersections and network links. Network links are made up of multiple zones of the same size. AGVs move from one zone to the next along these links. The movement of AGVs is governed by the end control rule, which dictates that a transporter releases its current zone at the end of its movement to the next zone. This rule ensures that multiple AGVs can travel on the same network link but not in the same zone. At GMH, the safety

Table 1.3: Model Objectives

Component	Details
Organizational Aim	
Aim 1	Reduce congestion on the mezzanine floor at GMH.
Modeling Objectives	
Objective 1	Determine whether swapping paths of clean and soiled case carts reduces congestion in the mezzanine floor, reduces trip time for clean and soiled case cart deliveries, reduces task completion time and is feasible to implement.
Objective 2	Determine the optimal number of AGVs to use to deliver clean surgical case carts from MD to CCSA by 7 pm daily. The second objective only needs be considered if the aim cannot be met by swapping of the paths of clean and soiled case carts.
Constraints	
Budget	The budget is limited.
AGV Guidepaths	The direction of AGV paths cannot be changed. AGV must move on preexisting paths.
AGV Parameters	AGV parameters cannot be changed.
Overall Process	The current material handling process cannot be altered. This processes ensures safety (decontamination) of materials.
General Project Objectives	
Flexibility	Limited flexibility since extensive model changes beyond changes to the data are not expected.
Run-Speed	A reasonably small running time is important due to the large number of experiments conducted.
Visual Display	A simple 2D animation is sufficient. The team members of PSD are familiar with the movement of surgical carts. Thus, animation is required verify that the model behaves corresponds to the real system.
Ease-of-use	The model is for use by the modeler thus interactive features are not required to be included.

distance between the AGVs, which is enforced at all times, is 3 feet. To ensure that AGVs maintain this distance in the model, we set a zone length of 3 feet on every network link in the model. When an AGV comes to a halt to maintain sufficient follow-up distance or yield to another AGV, it decelerates, temporarily stops, and accelerates again. We model all these phases of movement on the basis of the specifications provided in the AGV system handbook. Thus, micro-stoppages due to AGV traffic are accurately modeled in the simulations.

Destinations such as the CSSD and the MD are modeled using intersections on the network. Each destination is modeled as the last intersection on a path. The network links that connect to these destinations are modeled as bidirectional links. These links have a capacity limit of up to 1 AGV, which means that up to 1 AGV can travel to or from any destination on the corresponding link. At every intersection, the first come, first served rule is followed to determine the right-of-way for AGVs, unless another priority rule applies. If an AGV already has control of an intersection, another AGV must wait to use the intersection until the AGV leaves.

Some of the elevators can accommodate only one AGV and others can accommodate up to 2. In the latter case, if there is already an AGV in the elevator, the elevator waits for the next AGV if it is already at the preceding intersection. Otherwise, the elevator moves a single AGV. The detents on the second floor have limited capacity; therefore, when they are full, the AGVs are not allowed to enter the elevator to move to the second floor. AGVs leaving departments or elevators have a higher priority to seize intersections than AGVs entering departments or elevators. After completing a task, an AGV is assigned to the next request in the queue. If the queue is empty, the AGVs are moved to a parking area.

Movement of clean case carts Each surgical case is modeled as an entity. Every day, the first entity, whose type is *clean*, is created at 3 pm. Carts and employees loading soft goods into a cart are modeled as resources of the picking process. The distribution of the time required for the picking process is determined to be triangular, TRIA(2,3,5) minutes, based on data collected through a time study. Once the soft goods are loaded into the cart, an AGV is requested. An available AGV closest to the MD is assigned to the case cart. The AGV travels to the pickup location following the rules that govern the movement of the AGV. After picking up a case cart, the AGV travels to elevator J using "*Path of AGV with Clean Cart*", as shown in Figure 1.1. A predefined look-ahead stop is modeled before the intersection, "*Intersection J*", in the corridor between CSSD and elevator J. At this stop, the availability of intersection J and elevator J is checked, as well as the capacity to accommodate a vehicle at the CCSA, the destination. If these conditions are satisfied, the AGV seizes intersection J and advances to its final destination. Otherwise, the AGV is put on hold at the predefined stop until these conditions are satisfied. The clean case cart is dropped off at the CCSA. Details of the elevator logic in our simulation models are described in Figure 1.3. Clean Cart movements are completed by 7 pm every day and by midnight if there is a need for overtime. Thus, clean cart movements on different days do not interfere.

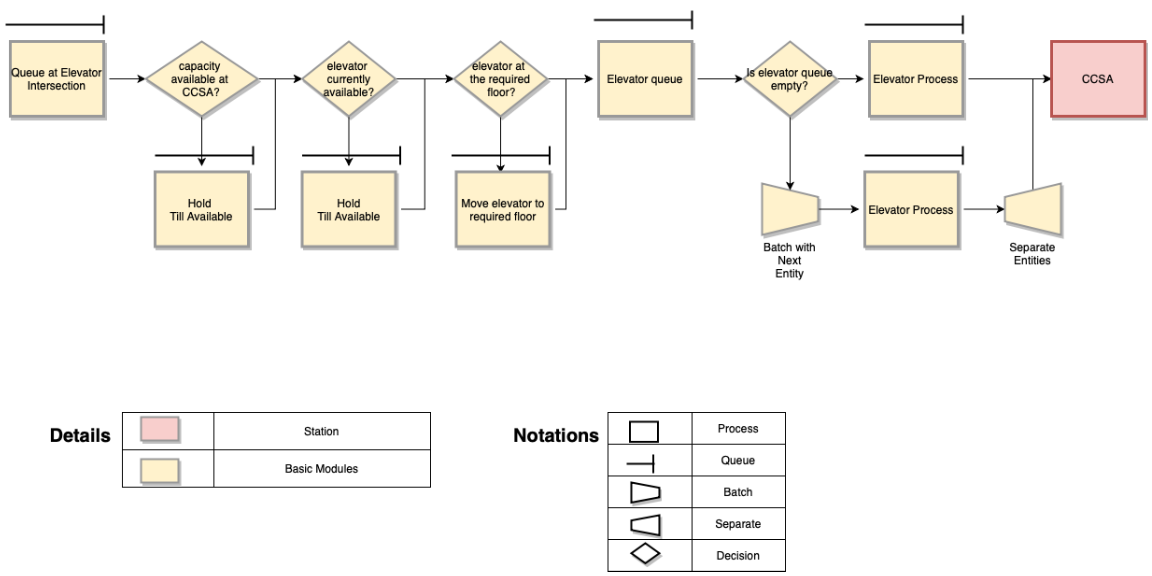


Figure 1.3: Flowchart of Elevator Logic

Movement of soiled case carts At 8 am the next day, the carts are released from the CCSA and moved to the ORs to prepare for surgery. The carts are released from surgeries based on the discrete distribution of release times (see Table 26 in the Appendix). After being released from the OR, a new entity type, called *soiled*, is assigned to the case carts. The soiled carts are moved to a location near the G and K elevators on the second floor, *Soiled Cart Storage Area* (SCSA). A request for an AGV is submitted and an available AGV closest to the SCSA is assigned to the case cart. The soiled carts are then transported to the CSSD along the "*Path of AGV with Soiled Cart*" shown in Figure 1.1. Soiled case cart movements start at 8 a.m. and continue throughout the day according to the surgery schedule. These movements end at 8 am the next day. Thus, cart movements in a soiled case on different days do not interfere. As a result, there is no accumulation of workload in the system. Given these characteristics, the only initialization effect in the model is due to the lack of dirty cart movement on the very first day of each replication. Observations show that this effect is minimal and does not affect the simulation results.

Movement of washed case carts The cart-washer is modeled as a resource with a fixed cycle time of 15 minutes. After an AGV drops a cart at the CSSD, the soiled instruments are separated from the cart, and the cart is loaded into the cart washer. After cleaning the cart, the closest available AGV picks up the cart and transports it to the MD along "*Path of AGV with Washed Cart*". The carts are subject to an additional 30 minutes of drying time before being released from the associated surgical case.

Simulation flowchart, assumptions, and simplifications Figure 1.4 presents the flow chart of material handling activities at GMH. Notice that we only focus on the movement of AGVs carrying surgical carts. Other AGV movements are absent in the simulations, so the movement of surgical carts is analyzed in isolation. The model we propose produces accurate results, despite this simplification because (*i*) the AGVs that carry other types of carts do not use the same paths; (*ii*) GMH allocates a fixed number of AGVs to surgical cart movements, and the availability of AGVs is not affected by the demand generated for AGVs by other material handling activities. Other assumptions made to meet the general objectives of the model without significantly affecting the results are summarized in Table 1.4. These assumptions and simplifications were discussed with PSD employees before the simulation model was developed.

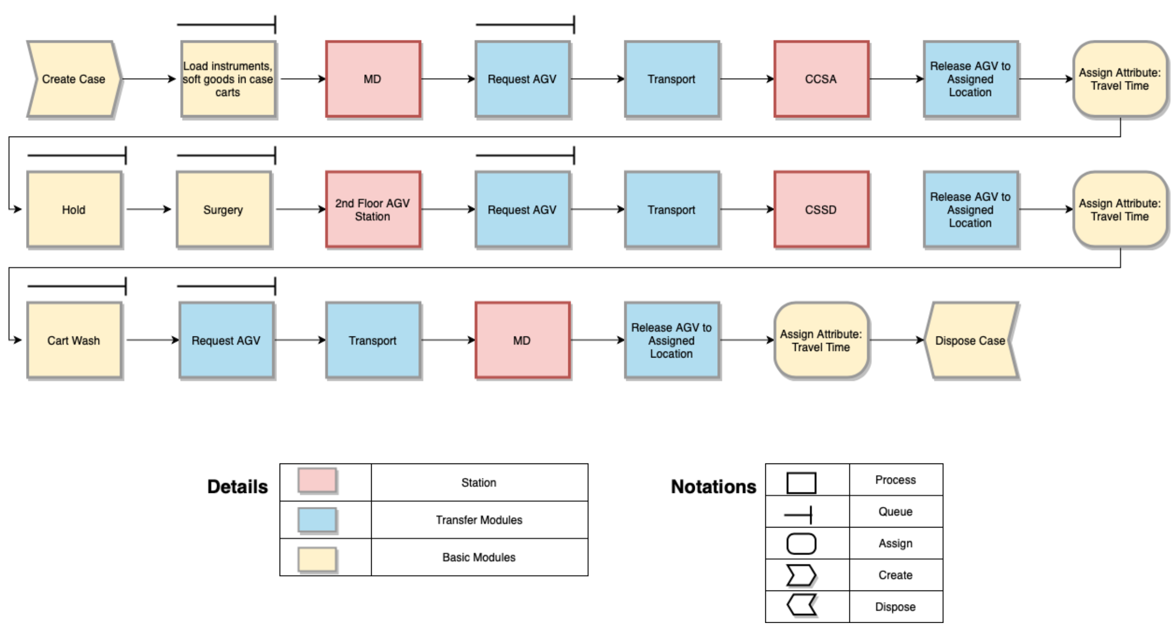


Figure 1.4: Flow Chart of Simulation Model

Table 1.4: Model Simplifications and Assumptions

Component	Details
Model Assumptions	
Material	Carts are always available for picking process.
Ors	Ors are always available.
Staff	Staff is always available.
Shift	No work is carried out during off-shift hours.
AGV Breakdowns	AGV breakdowns are rare and have little impact on AGV availability.
Cart washer Breakdowns	Cart washer breakdowns are rare and have little impact on cart availability.
Elevator Breakdowns	Elevator breakdowns are rare and have little impact on elevator availability.
Model Simplifications	
Other services	Other services that use AGVs do not impact the delivery of surgical case carts.
AGV Availability	At most 11 AGVs are available for the delivery of surgical carts.

1.4.2 Simulation Experiments

The simulation model is developed using ARENA© simulation software by Rockwell Automation. We present three sets of experiments. The first set of experiments was run to validate our model. The second and third sets of experiments were run to answer the two research questions we introduced. In all experiments, each day begins at 8 am and ends at 8 am the next day, which corresponds to the actual operating hours of the CSSD, where clean surgical carts are loaded and soiled case carts are cleaned. As pointed out before, the movements of clean and soiled case carts are initiated and completed within this 24-hour window, and the system does not reach a steady state as days go on. Despite this, we vary the replication length depending on the purpose of the experiment. For example, we generate the scenarios in our simulation-optimization experiments in Section 5.3.2 based on average performance measures. This necessitates running our experiments over a longer period of time, since our statistical analysis shows that days of the week differ in terms of case volumes. In other experiments, we used historical case volume data as input, which dictates the replication length. Additional information on the input data and the replication length for each set of experiments is reported in the corresponding subsections. In all experiments, we conducted 30 replications since further increasing the number of replications did not yield statistically different model outputs. Table 1.5 and Tables 4 and 5 in the Appendix provide details about the simulation model.

Table 1.5: Run-setup Parameters

Run-setup parameters	**Description**	**Run-setup parameters**	**Description**
No. of replications	30	Warm-up period	0
Base time units	Minutes	Statistics collection	Continuous

1.4.3 Validation

To ensure that the model presented in the previous section accurately reflects the logic and business rules of the real system, a statistical comparison of the current system with simulation model M is performed based on the travel times of clean and soiled case carts. Model M uses 11 AGVs, which is the same number of AGVs as the hospital currently uses every day for the movement of case carts. The simulation was run with the input data estimated for each day of the week (see 1.2) and the output of 30 replications for each day was compared to historical data. We conduct hypothesis tests to compare the average travel times in the data generated by model M and the average travel times obtained from the data of the AGV system in GMH. The null hypothesis is $H_0 : \mu_{MC} - \mu_{DC} = 0$ ($H_0 : \mu_{MS} - \mu_{DS} = 0$), and the Alternative hypothesis $H_1 : \mu_{MC} - \mu_{DC} \neq 0$ ($H_1 : \mu_{MS} - \mu_{DS} \neq 0$), where, $\mu_{MC}(\mu_{MS})$ is the average travel time of model M for clean (soiled) case carts on each day of the week. The average travel time obtained from the data for clean (soiled) case carts is represented by $\mu_{DC}(\mu_{DS})$. We used a two-sample t-test at a 95% confidence level to determine whether the travel times in model M and the data are significantly different from each other for clean and soiled case carts. Tables 1.6 and 1.7 summarize the results of the t-tests. Based on the confidence intervals, it is concluded that the difference between average travel times is not statistically significant, and thus the simulation model presented here is valid. These results were also verified by a PSD team at GMH.

Table 1.6: Model Validation: Travel Times of Clean Case Carts

	Current System (Data)			Model M		
Weekday	Mean	St. Dev.	Confidence Interval	Mean	St. Dev.	Confidence Interval
Monday	9.9533	3.2448	(6.9524,12.954)	9.9386	0.27121	(9.8374,10.040)
Tuesday	9.0296	2.8029	(6.4373,11.622)	10.003	0.2654	(9.9035,10.102)
Wednesday	9.3141	1.3458	(8.0695,10.559)	9.5709	0.29505	(9.4607,9.6811)
Thursday	9.2071	1.6106	(7.8607,10.554)	9.8108	0.29515	(9.7006,9.9210)
Friday	9.8126	2.684	(7.3303,12.295)	9.6367	0.36287	(9.5012,9.7722)

Research Question 1: Can performance measures, such as travel time and task completion time for AGVs, be improved after a redesign of AGV pathways at GMH?

Table 1.7: Model Validation: Travel Times of Soiled Case Carts

	Current System (Data)			Model M		
Weekday	Mean	St. Dev.	Confidence Interval	Mean	St. Dev.	Confidence Interval
Monday	6.841	1.5052	(5.4490,8.2330)	6.1604	0.16131	(6.1001,6.2206)
Tuesday	6.5821	1.0582	(5.6035,7.5608)	6.1847	0.18719	(6.1148,6.2546)
Wednesday	6.0904	0.62017	(5.5169,6.6640)	6.2396	0.10024	(6.2022,6.2771)
Thursday	6.3501	0.69885	(5.7659,6.9344)	6.3058	0.18458	(6.2368,6.3747)
Friday	6.1281	0.60938	(5.6187,6.6376)	6.1994	0.11993	(6.1546,6.2441)

Swapping the role of elevator J with elevators G and K would eliminate shared paths. However, there is a trade-off between congestion and travel distances, since the new design modifies the paths for clean and soiled carts. We have modeled and simulated this design change to evaluate the resulting trade-off. Figure 1.5 presents alternative AGV routes for clean and soiled case carts when the elevators are swapped. In this case, AGVs that follow the new clean case cart route take elevators G or K, and drop the clean case carts in the current system?s SCSA. After surgery, the soiled case carts are stored at the CCSA. AGVs carrying soiled case carts take elevator J and travel through intersection J to the sterilization area, CSSD. Model S is built to capture these changes.

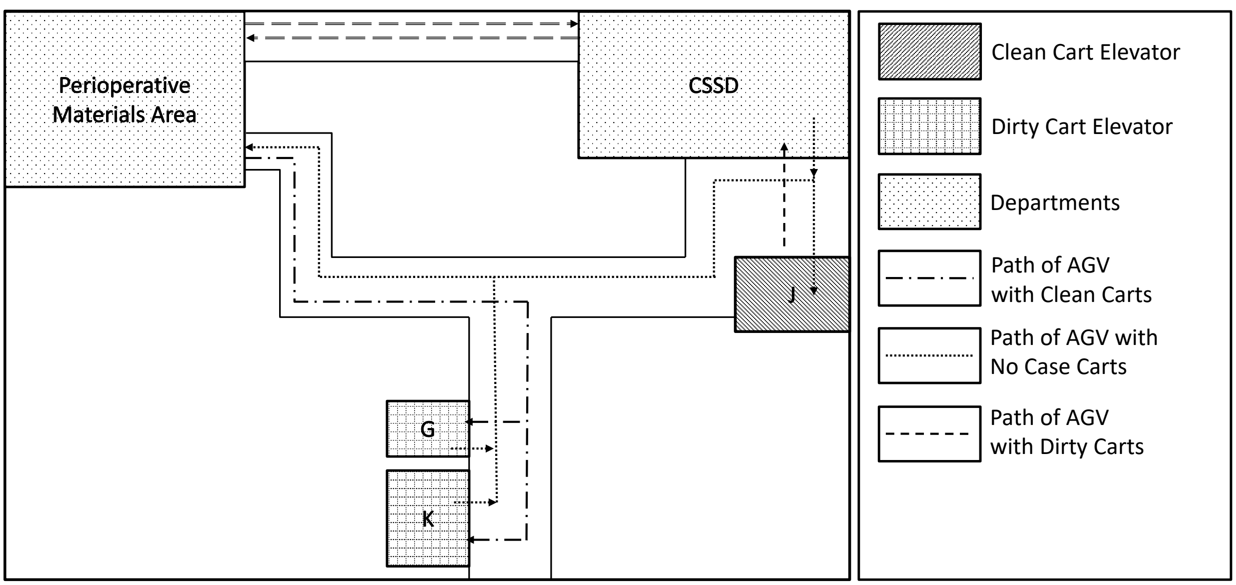


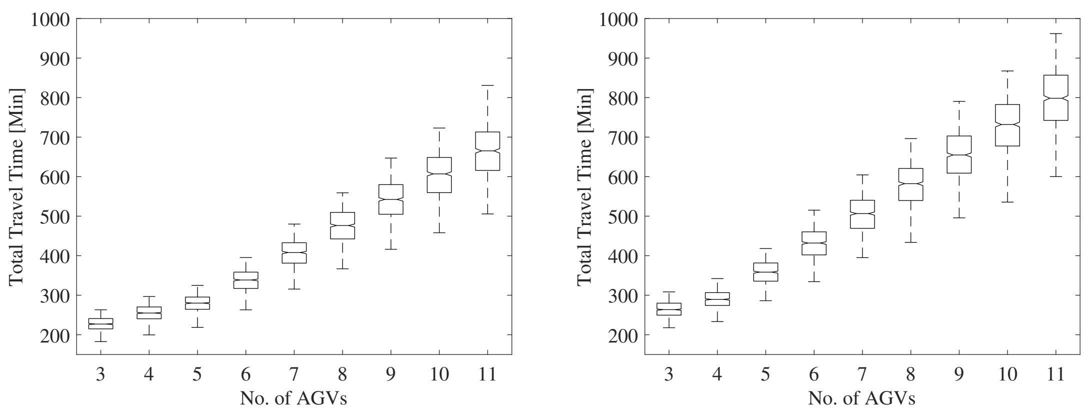
Figure 1.5: AGV System with Swapped Elevators (S)

All parameters and components of the model other than the AGV pathways are the same in models M and S. A set of simulation experiments was run by varying the number of AGVs in each model from 3 to 11, increasing with increments of 1. We used 36 weeks of historical data on case volumes as input and used common random numbers in the sampling of all other input data in Models M and S. The models are compared based on the average travel time and the *task completion time* (T_c) which is defined as the difference between the time the first clean case cart was picked up and the time when the last clean case cart of the day was dropped. A sensitivity analysis is performed to understand the impacts of the proposed changes on these measures. The results of this analysis are summarized in the next three subsections.

Results of the Sensitivity Analysis: Travel Times of Clean Cart Movement

Figure 1.6 summarizes the results of the sensitivity analysis for clean case carts. Figures 1.6a and 1.6b show the box plots of the total daily travel times for models M and S, respectively. We observe that, in model M, the total travel time increases with the number of AGVs. Since every AGV travels exactly the same distance, the increase in travel time is due to waiting in traffic. Traffic congestion increases with the number of AGVs in the system. Similarly, travel time is sensitive to the number of AGVs used in model S (see Figure 1.6b).

Figure 1.6: Sensitivity Analysis of Clean Case Carts: Model M vs Model S



(a) Total Travel Time Model M

(b) Total Travel Time Model S

We conduct an hypothesis test to compare the average travel times of model M and model S for clean case carts. Table 1.8 shows the descriptive statistics for each model. The null hypothesis is $H_0 : \mu_{dC} = 0$, and the alternative hypothesis is $H_1 : \mu_{dC} < 0$, where, μ_{dC} represents the difference in the average travel times of models M and S for clean case carts. We used a paired t-test at a 95% confidence level to determine whether the average travel time in model M is significantly smaller than the average travel time in model S. Table 1.9 shows the results of the t-test when the number of AGVs in both systems is 11. The difference in average travel times is statistically significant, and the average travel time for clean case carts is longer due to longer travel distances on the second floor if the elevators are swapped.

Table 1.8: Clean Case Cart Movement: Model M vs Model S

		Travel Time		
Systems	**N**	**Average**	**St.Dev**	**SE Average**
Model M	30	8.216	0.029	0.005
Model S	30	11.159	0.029	0.005

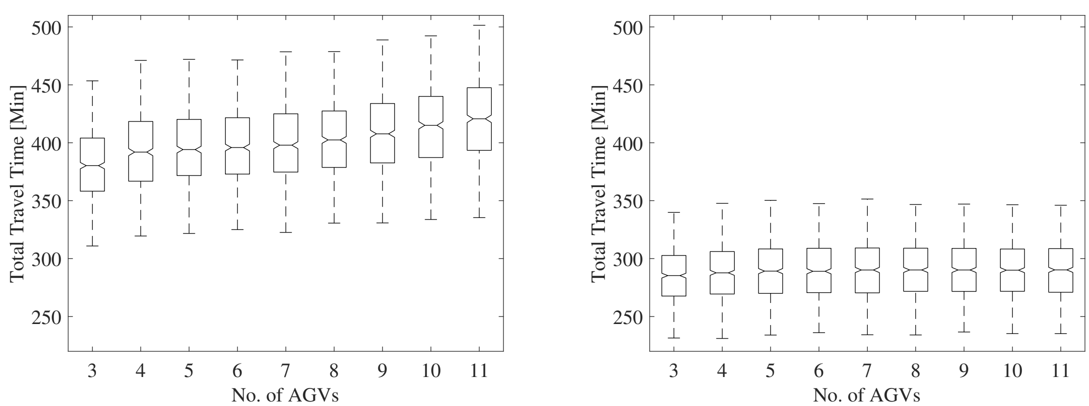
Table 1.9: Estimation for Paired Difference

Average	**StDev**	**SE Average**	**95% Upper Bound for μ_d**
-2.942	0.026	0.004	-2.935

Results of the Sensitivity Analysis: Travel times of the soiled cart movement

Figure 1.7 summarizes the results of the sensitivity analysis for soiled case carts. Figures 1.7a and 1.7b present box plots of the total daily travel times for models M and S respectively. In model M, an increase in the number of AGVs leads to longer travel times for soiled case carts due to congestion. This is mainly because AGVs with clean and soiled carts share paths. The travel times of the soiled carts in model S are not affected by changes to the number of AGVs.

Figure 1.7: Sensitivity Analysis Soiled of Case Carts: Model M vs Model S



(a) Total Travel Time Model M

(b) Total Travel Time Model S

We conduct an hypothesis test to compare the average travel times of model M and model S for soiled case carts. Table 1.10 shows the descriptive statistics for each model. The null hypothesis is $H_0 : \mu_{dS} = 0$, and the alternative hypothesis is $H_1 : \mu_{dS} < 0$, where, μ_{dS} represents the difference in the average travel times of the models M and S for the soiled carts. We used a paired t-test at a 95% confidence level to determine whether the average travel time in model M is significantly shorter than the average travel time in model S. Table 1.11 shows the results of the t-test when the number of AGVs in both systems is 11. The difference in average is statistically significant and the travel time for the soiled case carts is shorter if the elevators are swapped.

Table 1.10: Soiled Case Cart Movement: Model M vs Model S

		Travel Time		
Systems	**N**	**Average**	**St.Dev**	**SE Average**
Model M	30	6.217	0.008	0.001
Model S	30	4.266	0.003	0.001

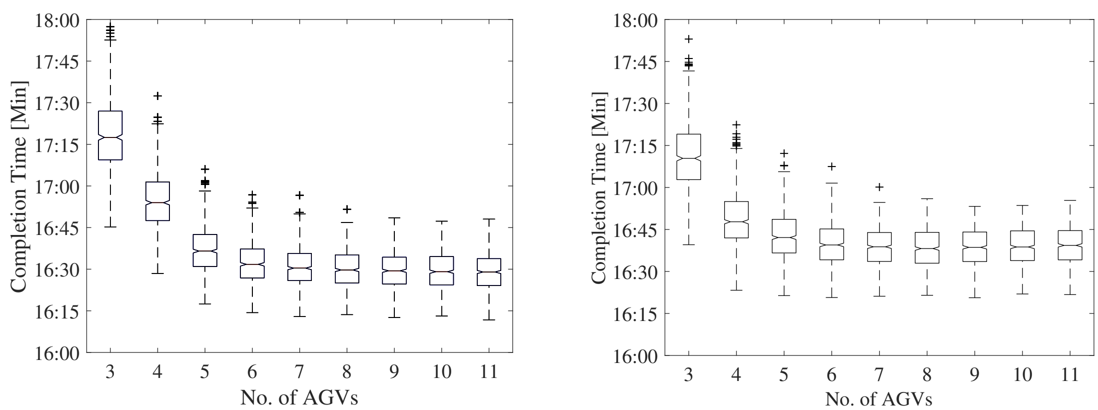
Table 1.11: Estimation for Paired Difference

Average	**StDev**	**SE Average**	**95% Upper Bound for μ_d**
1.951	0.009	0.002	1.954

Results of the Sensitivity Analysis: Task Completion Times

It is important to know the impact on task completion time if the hospital decides to use the same number of AGVs at all times. Figures 1.8a and 1.8b present the task completion time for models M and S. Task completion time decreases as the number of AGVs increases. A significant reduction in task completion times can be observed in both systems if 6 or 7 AGVs are used as opposed to using fewer AGVs. However, increasing the number of AGVs beyond 6 or 7 does not have a significant impact on task completion time.

Figure 1.8: Sensitivity Analysis Task Completion Times: Model M vs Model S



(a) Task Completion Time Model M (b) Task Completion Time Model S

We conduct an hypothesis test to compare the task completion times of model M and model S. Table 1.12 shows the descriptive statistics for each model. The null hypothesis is $H_0 : \mu_d = 0$, and the alternative hypothesis is $H_1 : \mu_d < 0$, where, μ_d represents the difference of average task completion times of models M and S. We used a paired t-test at a 95% confidence level to determine whether the average task completion time in model M is significantly shorter than the average task completion time in model S. Table 1.13 shows the results of the t-test when the number of AGVs in both systems is 11. The difference in average is statistically significant and the task completion time is shorter if the elevators are swapped.

Table 1.12: Clean Case Cart Movement: Model M vs Model S

		Task Completion Time		
Systems	**N**	**Average**	**St.Dev**	**SE Average**
Model M	30	242.832	1.612	0.294
Model S	30	145.141	1.891	0.345

Table 1.13: Estimation for Paired Difference

Average	**StDev**	**SE Average**	**95% Upper Bound for μ_d**
97.691	1.701	0.311	98.219

Average travel time is expected to increase with the number of AGVs in the system due to congestion. This is a well-known result and has been briefly discussed in [153] and [84]. However, it is not easy for the decision maker to estimate: (i) What is the marginal decrease/increase in travel time when the fleet size is reduced by 1 unit? (ii) What is the marginal decrease/increase in task completion time when the fleet size is reduced by 1 unit? (iii) What is the minimum number of AGVs needed to complete all tasks in the required time? This additional information helps decision makers to make better informed decisions.

Research Question 2: Do performance measures, such as travel time and task completion time, improve when the number of AGVs used daily is controlled by the volume of surgical cases?

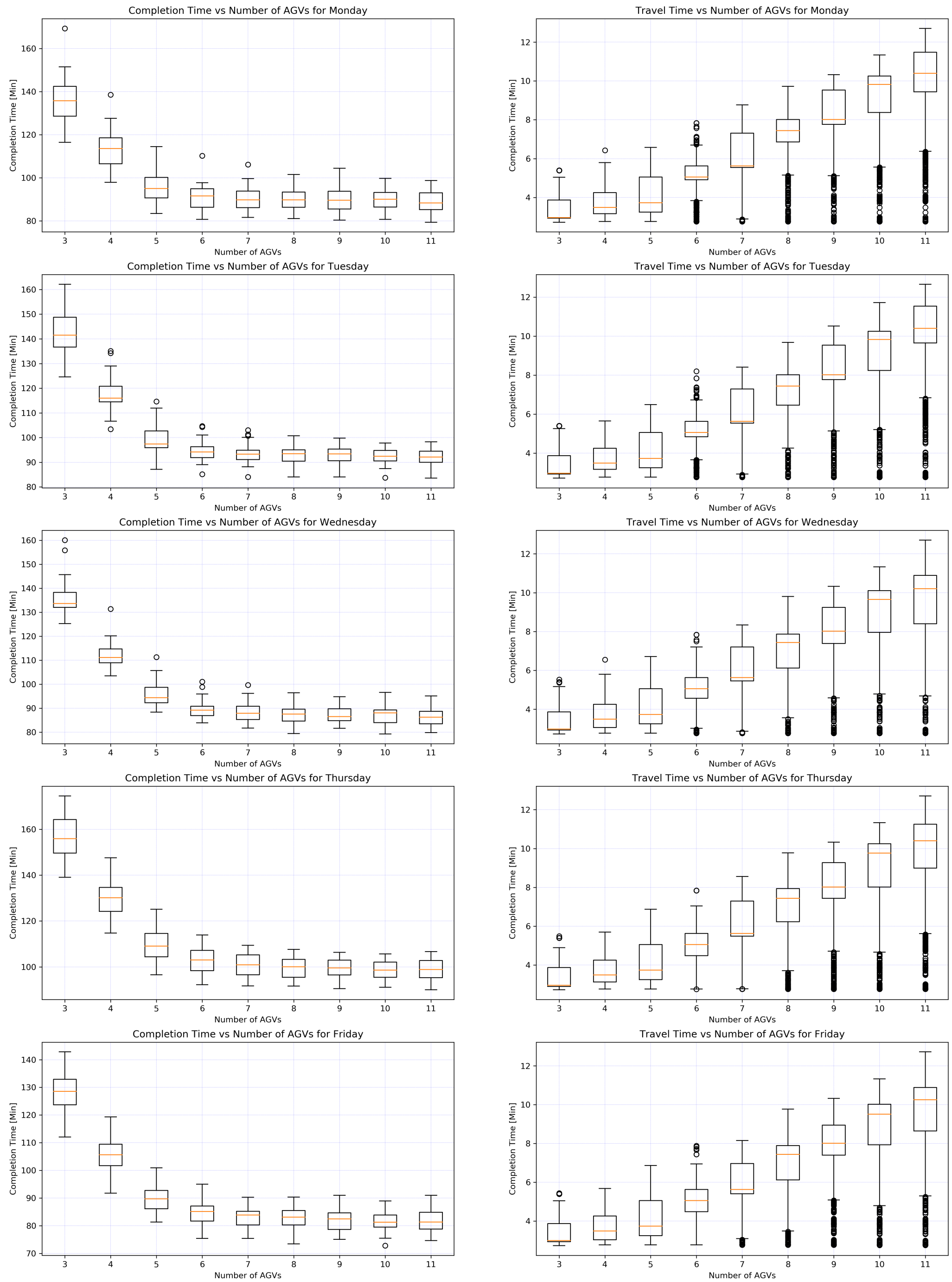
Data analysis shows that the volume of surgical cases follows a distribution whose mean value changes depending on the day of the week. Therefore, the hospital would expect to use a different number of AGVs to deliver surgical case carts on each day, instead of using a fixed number of AGVs every day. This approach could lead to improved AGV utilization, reduced congestion, and shorter travel times. Moreover, such control of the number of AGVs that are in use is practical, since it can be implemented with minor changes in the control system or simply limiting the number of AGVs requested. For example, a new AGV may not be requested until one of the active AGVs has completed its task. Thus, the case carts could be held at detents until one of the AGVs that are in use is available again.

We propose that only $k \leq 11$ AGVs should be used on a particular day, where k is adjusted according to the day of the week. Note that the hospital currently uses 1 AGV every day. To find the number of AGVs $k \leq 11$ must be used, we create a separate simulation model for each weekday. The number of surgical cases, that is, clean case carts to be carried for each day, was generated based on the distribution given in Table 1.2. This number also determines the number of soiled case carts to be delivered the next day. Using these distributions, we simulate each weekday for 30 replications. For each model, the number of AGV is fixed for the entire simulation run. We repeat the experiments by varying the number of AGVs from 3 to 11 AGVs for each weekday. Based on the results, the AGV fleet sizes that yield the lowest average travel time and the lowest task completion time can be selected on each day.

Results of the Operational Fleet Sizing Experiments

The results of our experiments are shown in Figure 1.9. The results indicate that varying the number of AGVs used on each day of the week is beneficial for the hospital. For example, 6 AGVs can be used on Mondays to achieve a shorter task completion time. To achieve similar task completion times on Tuesdays, the hospital must choose more than 10 AGVs. However, increasing the number of AGVs increases task completion time and, thereby, congestion. Thus, the number of AGVs to be chosen for each day is different depending on the management objective.

Figure 1.9: Results of Operational Fleet Sizing Experiments



Operational Fleet Sizing Experiments via Simulation-Optimization

In the previous section, we presented a method to choose the number of AGVs for each day. The number of AGVs to be used for each day depends on the objective of management. However, it is also important for the decision maker to see the overall impact of choosing a fleet of AGVs that meets the required service level. Therefore, to generate alternative scenarios that balance congestion levels and task completion times, we propose two experiments using ARENA OptQuest. Both experiments focus on optimizing AGV movements on "*Path of AGVs with Clean Cart"* since analysis shows the movement of clean carts causes congestion on the mezzanine floor. For both experiments, the decision variable is the *number of AGVs to be used on a particular day of the week*. To obtain average performance measures over time, we simulate each scenario for 30 days and use 30 replications at each point. For the number of AGVs used on day d, k_d, the search interval for controls is specified as $3 \leq k_d \leq 11$.

In experiment 1, the objective is to minimize the total travel time each day. It is also important that the movements of all clean case carts are completed by 7 pm. To ensure that, in experiment 1, a task completion time constraint $T_c \leq 200$ minutes is added. In experiment 2, the objective is to minimize the sum of task completion times over a replication. Tables 1.14 and 1.15 summarize the results of the OptQuest experiments. These tables present solutions that satisfy the following three conditions: *(i)* the total number of AGVs used in a day is less than or equal to 11; *(ii)* the average travel time per AGV is less than or equal to the average travel time observed from the data; and *(iii)* the total completion time is not later than 5:05 pm. These criteria identify solutions that could potentially be adopted by GMH. Each solution presents the minimum, maximum, and average travel time for each AGV; the task completion time; and the number of AGVs used each day.

The results in Table 1.14 suggest the use of fewer AGVs than the current practice in GMH because the objective of experiment 1 is to minimize the total travel time. Utilizing fewer AGVs leads to reduced congestion, as evidenced by the average travel time and the corresponding range of travel time, which is narrower. On the contrary, when the objective is to minimize the task completion time, the simulation experiments suggest using relatively more AGVs, as can be seen in the results in Table 1.15. This increase in the number of AGVs leads to congestion, evidenced by the average travel time and the corresponding range of travel time, which is wider compared to the results in Table 1.14. However, the completion time is affected by both the travel time and the waiting time. As a result, the task completion time is shorter for solutions with a higher number of AGVs available because the case carts spend less time waiting for AGVs. The solutions of OptQuest use a different number of AGVs each day of the week, which is different from the current practice in GMH. Once again, the experimental results suggest that on days with a lower volume of cases, fewer AGVs should be used than on days with a higher volume of cases.

Table 1.14: Results of Experiment 1: Minimize Total Travel Time Per Day

	Travel Times [Min]			Task Completion Time	Number of AGVs				
Solution	Min	Max	Average	Average	M	T	W	Th	F
1	2.64	6.47	3.43	5:03:01 PM	3	3	3	4	4
2	2.64	6.78	3.58	4:55:51 PM	3	4	4	3	5
3	2.64	9.72	5.37	4:41:20 PM	3	4	8	7	8
4	2.64	12.1	5.35	4:37:50 PM	4	4	5	10	7
5	2.64	6.78	3.84	4:43:25 PM	5	4	4	4	5
6	2.64	7.93	4.33	4:34:51 PM	5	4	6	5	6

Table 1.15: Results of Experiment 2: Minimize the Task Completion

	Travel Times [Min]			Task Completion Time	Number of AGVs				
Solution	Min	Max	Average	Average	M	T	W	Th	F
1	2.64	12.1	7.27	4:28:46 PM	8	7	8	10	8
2	2.64	12.1	7.96	4:28:30 PM	10	7	10	10	8
3	2.64	13.64	8.68	4:28:01 PM	11	6	11	11	11
4	2.64	13.64	8.57	4:28:04 PM	11	7	10	10	11
5	2.64	13.64	8.9	4:27:49 PM	11	7	11	11	11

1.4.4 Implementation

To further evaluate the impact of the proposed changes on AGV utilization, travel time, task completion time, and congestion, the solutions obtained from the simulation experiments were implemented using the following approaches: First, a simulation study was conducted using real-life data from GMH regarding the total number of surgical cases performed each day of the week, from January 1, 2018, through September 11, 2018. This 36-weeks worth of data allowed for a thorough statistical analysis of the results. A fleet of AGVs was selected to deliver surgical carts each week of this period. In these experiments, the less conservative solutions presented in Table 1.15 were used instead of the solutions presented in Table 1.14 because the GMH staff expressed concerns about the potential delays that can result from significantly reducing the number of AGVs used each day (from 11 to 3, 4, or 5). On the basis of the results, a further analysis is presented for the solution that was selected for a pilot study. Section 1.4.4 summarizes the results of this analysis. A short pilot study was then conducted at GMH. This study was only one week long due to the additional resources needed for implementation. Section 1.4.4 summarizes the results of this study. Section 1.4.5 presents the managerial insights revealed by the simulation experiments and the pilot study.

Implementation via Simulation

To further evaluate the scenarios generated in Section 1.4.3, we ran an additonal set of simulation experiments. The actual volume of cases, which was collected by GMH for 36 weeks between January 1, 2018, and September 11, 2018 was used as input. Figure 1.10 summarizes this data. Table 1.16 and Figure 1.11 summarize the results of the simulation runs for the solutions obtained from experiment 2 (see Table 1.15).

Figure 1.10: Case Study Data: Volume of Surgical Cases per Month

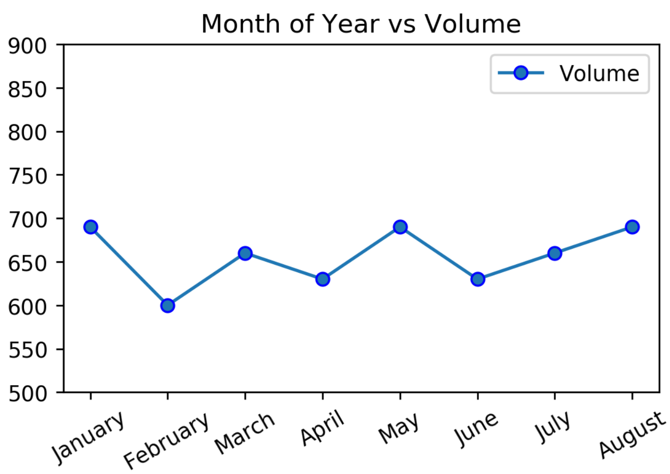
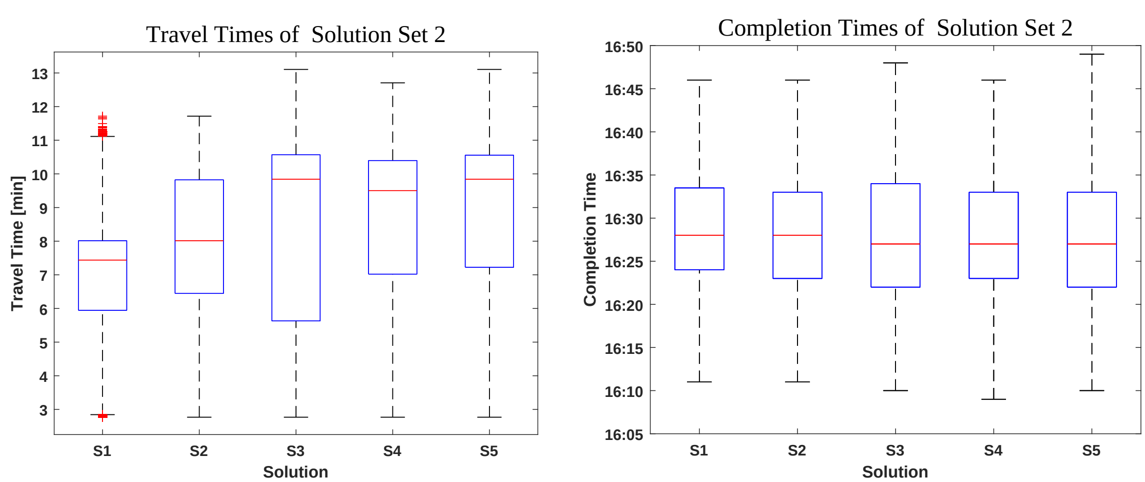


Figure 1.11: Evaluation of Policies from Experiment 2: Minimize Task Completion Time



(a) Travel Time

(b) Task Completion Time

The simulation results confirm our earlier observations with respect to the simulation-optimization results. The simulation results show that the solutions considered here yield comparable task completion times, which are, on average, around 4:30 pm, well before the target completion time determined by the hospital. On the other hand, the average travel time attained varies among the solutions considered. The robustness of the task completion times can be ex-

Table 1.16: Evaluation of Policies from Experiment 2: Minimize Task Completion Time

Solution	Travel Time [min]			Task Completion Time		
	Average	Confidence Interval	Std Dev	Average	Confidence Interval	Std Dev [min]
1	7.24	(7.23,7.25)	1.71	16:29:07	(16:28:47,16:29:27)	6.55
2	7.96	(7.93,7.95)	2.03	16:28:49	(16:28:49,16:29:09)	6.64
3	8.77	(8.75,8.79)	2.77	16:28:24	(16:28:01,16:28:46)	7.38
4	8.63	(8.61,8.64)	2.33	16:28:19	(16:27:57,16:28:40)	7.07
5	8.98	(8.97,9.00)	2.53	16:27:59	(16:27:59,16:28:21)	7.06

plained by the trade-off between travel times and the number of simultaneous trips possible (that is, the number of AGVs), which is again consistent with earlier results. Based on these results, we choose solution 1 from Table 1.15 for implementation, as this solution performs better than other solutions presented and uses fewer than 11 AGVs on all days. To evaluate the impact of this selected solution, we compare it with the current system in GMH. Tables 1.17 and 1.18 present the results of these simulation runs. Once again, we perform 30 replications in these experiments.

Table 1.17: Implementation via Simulation: Average Travel Times (in min)

Day	Simulation of Implemented Solution			Simulation of Current Practice		
	Avg. Travel Time	St.Dev	Confidence Int.	Avg. Travel Time	St. Dev	Confidence Int.
Monday	6.82	0.036	(6.810, 6.838)	9.61	0.057	(9.593, 9.636)
Tuesday	5.84	0.036	(5.827, 5.853)	9.61	0.072	(9.585, 9.639)
Wednesday	6.45	0.036	(6.434, 6.460)	8.96	0.048	(8.943, 8.979)
Thursday	7.96	0.057	(7.936, 7.978)	8.77	0.073	(8.738, 8.792)
Friday	6.55	0.041	(6.533, 6.563)	9.22	0.059	(9.198, 9.242)

Table 1.18: Implementation via Simulation: Average Completion Times (in min)

Day	Simulation of Implemented Solution			Simulation of Current Practice		
	Avg. Comp. Time	St.Dev	Confidence Int.	Avg. Comp. Time	St. Dev	Confidence Int.
Monday	119.37	2.436	(118.5, 120.3)	117.91	2.483	(116.98, 118.84)
Tuesday	107.54	1.829	(106.9, 108.2)	104.96	1.725	(104.32, 105.61)
Wednesday	176.31	7.175	(173.6, 179.0)	175.18	7.302	(172.45, 177.91)
Thursday	177.73	6.021	(175.5, 180.0)	177.41	6.019	(175.17, 179.66)
Friday	119.94	2.259	(119.1, 120.8)	118.52	2.281	(117.67, 119.37)

The simulation results summarized in Table 1.17 show that the average travel time in the proposed system significantly reduces the average travel times on all days of the week. Therefore, it is beneficial to *(i)* to vary the number of AGVs according to the day of the week (that is, based on the volume of the case) and *(ii)* to use fewer AGVs in general.

The proposed solution results in longer average completion times for each day of the week than current practice, but the differences in average completion times are relatively small and only significant on Monday, Tuesday, and Friday. Furthermore, the average completion time under the proposed solution is within the 200-minute period preferred by GMH on all days of the week.

It is important to understand the implications of our selected solution on the service level. Here, the level of service is measured by the fraction of days in which the task completion time exceeds 200 minutes. It was observed that out of 5,400 days, the simulated task completion time of the implemented solution exceeds the 200 minutes on 615 days, compared to 612 days in the simulated task completion times of current practice. From Table 1.19, we can see that on days when the completion time was exceeded, on average only 5% to 9% of the case carts were delivered after 200 minutes. For days when surgical cases were delivered after 200 minutes, Table 1.20 characterizes the average delay. It can be seen that on average, surgical cases were delivered late by 20 to 80 minutes. This confirms that the proposed solution achieves the same level of performance as is achieved by using 11 AGVs.

Table 1.19: Comparison of Two System: Fraction of Deliveries Delayed

	Simulation of Current Practice				**Simulation of Implemented Solution**			
WeekDay	**Average**	**Std. Dev.**	**Min**	**Max**	**Average**	**Std. Dev.**	**Min**	**Max**
Monday	0.05	0.032	0.021	0.168	0.048	0.031	0.021	0.168
Tuesday	0.041	0.017	0.021	0.074	0.042	0.018	0.021	0.074
Wednesday	0.061	0.038	0.019	0.178	0.062	0.038	0.019	0.178
Thursday	0.088	0.045	0.02	0.196	0.088	0.045	0.02	0.196
Friday	0.049	0.027	0.019	0.13	0.049	0.027	0.019	0.13

Table 1.21: Results of Pilot Study at GMH: Average Travel Times

Day	Week Before ($t-1$)		Treatment Week (t)		Week After ($t+1$)	
	Case Vol.	Avg. Travel Time	Case Vol.	Avg. Travel Time	Case Vol.	Avg. Travel Time
Monday	28	9.07	26	5.50	30	11.37
Tuesday	23	17.13	34	9.53	21	16.71
Wednesday	30	8.57	30	7.73	**14**	5.35
Thursday	24	16.79	26	8.54	25	9.35
Friday	30	10.37	32	8.00	22	9.26

Table 1.20: Comparison of Two System: Delays in Completion Time

	Simulation of Current Practice				Simulation of Implemented Solution			
WeekDay	**Average**	**Std. Dev.**	**Min**	**Max**	**Average**	**Std. Dev.**	**Min**	**Max**
Monday	24.187	20.8	1.65	98.238	22.963	21.095	1.979	100.792
Tuesday	21.564	14.418	2.253	47.583	21.359	14.327	1.978	48.727
Wednesday	67.293	98.706	1.65	541.485	67.689	98.225	1.65	540.08
Thursday	87.108	92.287	1.65	531.083	87.094	92.168	1.65	531.482
Friday	49.672	51.565	1.65	271.258	49.678	51.845	1.65	271.258

A Pilot Study at GMH

Our proposed system was piloted at GMH for one week. Thus, the analysis presented in this section is based on actual travel-time data obtained from physical implementation at GMH. During the pilot study, we visited GMH every day and collected data on AGV movement from 3:45 pm to 5 pm. Other data used in this section were obtained from the AGV control system. This implementation allowed us to evaluate how reducing the number of AGVs would impact congestion and travel time in the real system.

Table 1.21 presents the actual travel times of the AGVs during 3:45 pm to 5 pm each day of the treatment week (t), the week before ($t-1$), and the week after ($t+1$). We conducted a hypothesis test to compare the average travel times during weeks t and weeks $t-1$ and $t+1$ for clean case carts. The null hypothesis is $H_0 : \mu_t - \mu_{t-1} = 0$ ($H_0 : \mu_t - \mu_{t+1} = 0$), and the alternative hypothesis is $H_1 : \mu_t - \mu_{t-1} < 0$ ($H_1 : \mu_t - \mu_{t+1} < 0$), where, $\mu_t, \mu_{t-1}, \mu_{t+1}$ are the average travel times of weeks $t, t-1, t+1$ respectively. We used a two-sample t-test at a 95% confidence level to determine whether the average travel time is significantly shorter in week t than the travel times during week $t-1$ and $t+1$ for clean case carts. Table 1.22 provides the p-values of the tests performed.

Table 1.22: Results of Pilot Study at GMH: P-Values (Avg. Travel Times)

Treatment Period	**p-Values (Avg. Travel Times)**	
Day	**Week Before**	**Week After**
Monday	0.008	0.001
Tuesday	0.001	0.001
Wednesday	**0.262**	**0.999**
Thursday	0.001	**0.299**
Friday	0.012	**0.220**

The average travel time during the week t was lower than the average travel time during the week $t-1$ on all five days. This difference was statistically significant on Monday, Tuesday, Thursday, and Friday. On the other hand, the average travel time during the week t was lower than the average travel time during week $t+1$ on four days. This difference was statistically significant on Monday and Tuesday. The results show that the average travel time during week t was longer than during the week $t+1$ only on Wednesday, and the difference was statistically significant. This difference can be attributed to the fact that the number of cases on Wednesday in week $t+1$ was less than half the number of cases in the treatment week on the corresponding day. For very low case volumes, the material handling system is not significantly affected by congestion, and the use of more AGVs does not have serious adverse effects.

Table 1.23: Results of the Pilot Study at GMH: Standard Deviation of Travel Times

Treatment Period	**Std. Deviation**		
Day	**Week Before**	**Treatment Week**	**Week After**
Monday	6.97	1.72	3.82
Tuesday	9.67	3.34	9.02
Wednesday	6.06	3.46	**1.34**
Thursday	7.53	2.58	7.12
Friday	4.83	2.68	7.19

Table 1.23 presents the standard deviation of actual travel times during the weeks $t-1$, t, and $t+1$. We conducted a hypothesis test to compare these standard deviations for clean case carts. We tested the null hypothesis $H_0 : \sigma_t - \sigma_{t-1} = 0$ ($H_0 : \sigma_t - \sigma_{t+1} = 0$), and the alternative hypothesis $H_1 : \sigma_t - \sigma_{t-1} < 0$ ($H_1 : \sigma_t - \sigma_{t+1} < 0$), where, $\sigma_t, \sigma_{t-1}, \sigma_{t+1}$ are the standard deviations of the travel times during weeks $t, t-1, t+1$ respectively. We consider a significance level of 95%. Table 1.24 provides the p-values of the tests performed. The standard deviation of the travel times during

the week t is less than the standard deviation of travel times during the week $t-1$. This difference is statistically significant for all five days. On Monday, Tuesday, Thursday, and Friday, the standard deviation of travel time is lower during week t than during week $t+1$. This difference is statistically significant on Monday and Tuesday. Similarly to observations related to average travel time, the standard deviation of travel time on Wednesday of week $t+1$ is significantly lower than for week t.

Table 1.24: Results of the Pilot Study at GMH: P-Values (Std. Deviation of Travel Times)

Treatment Period	**P-Values (Std. Deviation of Travel Times)**	
Day	**Week Before**	**Week After**
Monday	0.001	0.001
Tuesday	0.001	0.001
Wednesday	0.018	**1.000**
Thursday	0.001	**0.071**
Friday	0.001	**0.088**

It is already established that longer travel times indicate longer wait times due to congestion. Similarly, the standard deviation of travel time is a measure of congestion in the system, i.e. a higher standard deviation, while the traveled distance is the same, indicates longer wait times due to congestion. The analysis of the results of the pilot study clearly shows that congestion was reduced by limiting the number of AGVs in the system, which led to reduced wait times and, consequently, to reduced travel times.

Limitations of pilot study: The data collected through the pilot study are not extensive due to the short implementation period. In addition to that, during the treatment week, the movement of the surgical carts began about 3:30 to 3:45 on 2 days, so only about 60% of the carts were delivered by 5 pm. At 5 pm, the AGVs were assigned to other tasks (e.g., delivery of dinner), so the rest of the carts were delivered later in the evening at about 9 pm, when the AGVs were available. Due to this lack of data, completion time is not reported.

1.4.5 Managerial Insights

Implications of Our Findings: This research was motivated by inefficiencies in the material handling system at GMH. During the afternoon hours of 3 to 5, AGVs with clean and dirty case carts used the same corridor, leading to increased congestion and longer travel times. GMH staff was interested in developing analytical solutions that would lead to reduced congestion in the main corridor of the mezzanine floor.

Research Question *(i)*: GMH staff suspected that a change of roles from elevator J to elevators G and K would reduce the congestion on the mezzanine floor. The research team conducted an extensive data analysis of the trips made by the AGVs to develop a simulation model in which the role of elevator J was swapped with elevators G and K. The results of the simulation model indicated a lower congestion on the mezzanine floor, supporting the intuition of the GMH staff. However, the simulation model also indicated longer overall travel times. In the current system, elevators G and K carry AGVs with soiled case carts, as well as trash and soiled linen carts, to the ground floor. If elevators are swapped, according to safety guidelines, soiled linen and trash carts must be delivered using different elevators. Several elevators were considered for these movements. However, the use of these alternative elevators increases the distance that AGVs travel to reach an appropriate detent area. It also increases the distance an employee travels to move soiled linen and trash carts to the new detent area. In addition, for the trash and soiled linen carts to use elevator J, a new AGV guide-path must be installed in front of the elevator on the ground floor. The installation of guide paths is expensive. Despite the benefit of reducing congestion on the mezzanine floor, the swapping of elevators was found to be costly and difficult to implement. Although the solution was not implemented, GMH staff found the results of our model beneficial, since they revealed tradeoffs and challenges that GMH staff had not foreseen.

Research Question *(ii)*: The results of the simulation-optimization model indicate that reducing the number of AGVs used each day and changing the number of AGVs based on the volume of cases would lead to reduced congestion, shorter travel times and shorter task completion time. The idea of reducing the number of AGVs used daily was received with doubt by the cart loaders because they were concerned that it would lead to delays in delivery time and consequently overtime work. Management supported the idea, but the implementation of the new system requires updates of the software that governs the movement of AGVs. These updates are completed by FMC-Technology at a fixed cost.

We note that, while the development of the proposed models was motivated by material handling activities at GMH, the approach we take for modeling the problem, analyzing the data, conducting sensitivity analysis and deriving managerial insights is generalizable and can help other hospitals identify opportunities to improve internal logistics. Finally, the managerial insights we provide attest to the value of using simulation-optimization. Intuition and expert opinions, while valuable, may not always yield the best available solutions to operational problems in a healthcare setting.

Limitations of Our Findings: A limitation of this study is that the model developed here focuses only on the movement of AGVs that deliver surgical case carts. This model can be extended to consider the movement of other AGVs. Increasing the scope of the model would result in a more accurate modeling of AGV availability, AGV traffic, and interactions between different services that use AGVs.

The volume of cases varies by day of the week. It also varies by week/month of the year. In Table 6 in the appendix, we report the fraction of case carts that were delivered after 200 minutes by month of the year. The average value of the fraction of case carts delivered late is between 4% - 8%. Table 7 in the Appendix compares the travel times of the simulation experiments of the implemented solution with the simulation of the current system. The average value of travel time is around 7 minutes for the solution we implemented for each month, while it is around 9 minutes if 11 AGVs are to be used. These values do not follow the small increases or decreases in the number of cases we see in Figure 1.10. Therefore, the results we present and the solutions we propose in this article are not affected by seasonal demand. However, this could be the case in other hospitals. Such seasonality effects require additional experiments to identify the correct system configuration during each demand season.

1.5 Integrated Material Handling and Inventory Management Processes:

Figure 1.12 describes a typical material handling process for the delivery of surgical case carts in a hospital. The process begins by creating a detailed schedule of surgeries. This schedule is prepared by the OR manager. On the basis of the schedule and preferences of the physicians, a list of instruments and soft goods is prepared and submitted to the Materials Department (MD). For each surgery, a clean case cart is loaded with the instruments, soft goods, and implants requested. These case carts are moved to pick-up/drop-off stations for AGVs to pick-up. The clean case carts are then moved from the MD to the storage area of the case cart (CCSA). At the CCSA, each case cart is inspected to ensure that it contains the required materials. The case carts are held at the CCSA until they are moved to the corresponding OR at the time of surgery. Case carts are delivered to ORs prior to surgeries. ORs are divided into separate cores on the basis of the specialties they serve. Specialty instruments and implants required for surgical cases, which are stored in the OR cores, are added to the case carts prior to surgery. After the surgery, instruments in the opened trays, instruments that were removed from peel packs, and case carts are considered soiled and should be decontaminated. The soiled carts and instruments are transported to the CSSD by the AGV. Clean instruments are returned to MD. Instruments and case carts are washed and sterilized at CSSD. Specialty surgical instruments are returned to the corresponding OR cores. This process ensures the availability of instruments prior to scheduled surgery. We use previously developed models. Thus, Figure 1.13 is presented again, indicating the locations of the departments and the paths traversed by the AGVs in GMH.

To reduce inventory cost, hospitals must coordinate inventory management and material handling decisions. This coordination becomes ever more important in the face of uncertainty. For example, if the duration of surgery and the travel time of the AGV are fixed, hospitals can calculate the necessary inventory levels with certainty and decide how

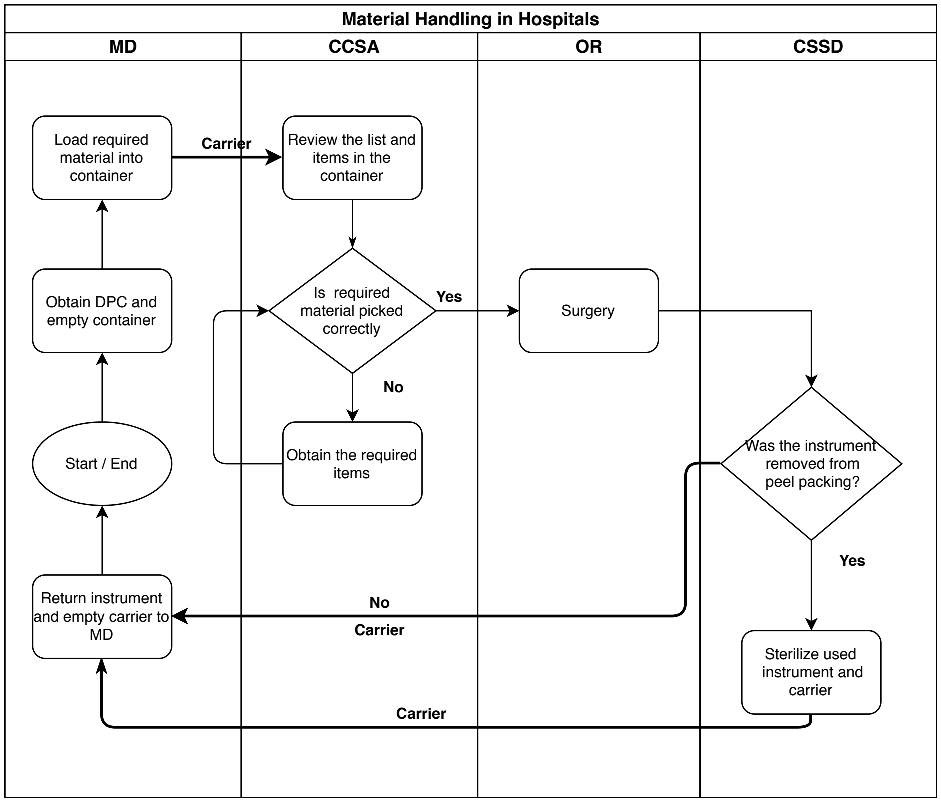


Figure 1.12: Material Handling Process

many instruments to loan or consign. However, to ensure a high level of service under uncertainty, many hospitals keep large inventories, loan instruments, and prepare/deliver case carts one day before surgery. In our partner hospital we observed that if the delivery of instruments from CSSD to OR was completed in a short time before the start of surgery, the instruments could be reused within the same day. This approach has the potential to lead to a reduction in the cost of using loaned or consigned instruments. This observation led to the development of the next proposed material handling process.

Experimental Setup: The movement of clean instruments to ORs and the movement of soiled instruments to the CSSD affect inventory availability and the start times of surgeries. This is the reason why some hospitals, such as GMH, prepare and deliver surgical case carts to the CCSA one day in advance. Such a practice ensures the availability of surgical instruments, but there are a number of inefficiencies regarding material handling and inventory management. For example, Table 1.25 summarizes the data for a period of 50 days obtained on the travel times of AGVs in the partner hospital. The data show that the average travel time and the corresponding standard deviation are highest during 3 pm to 6 pm. This is because clean surgical carts are being delivered to the CSSA for elective surgeries. The consequent increase in the number of AGV movements leads to congestion and, thus, longer travel times for every

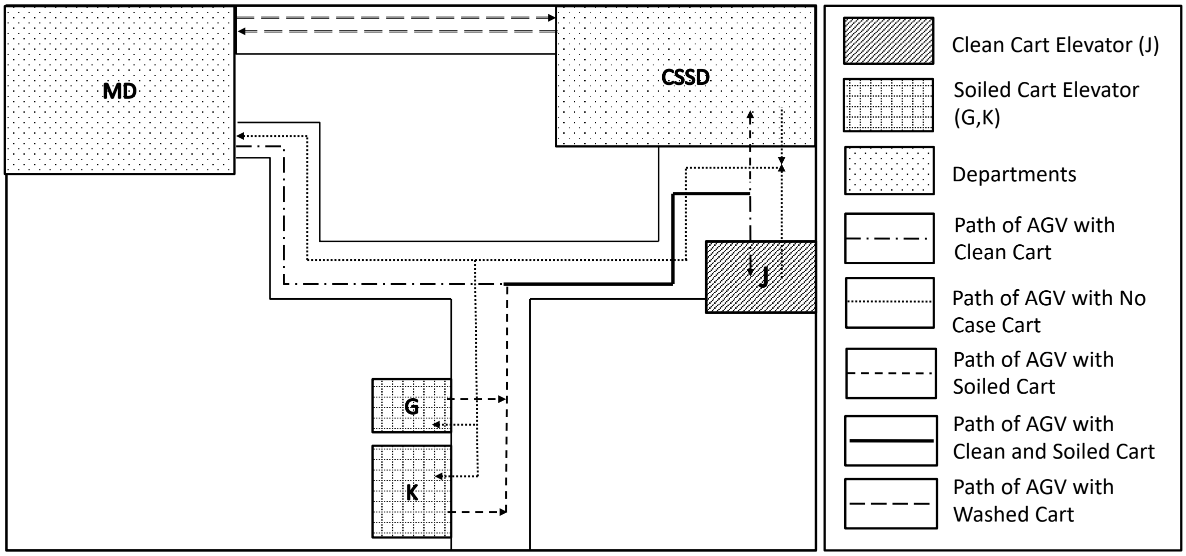


Figure 1.13: Map - GMH Floor Map

AGV that uses the same path. These delays lead to an increased inventory of instruments since an instrument cannot be reused in different surgeries scheduled on the same day. Table 1.26 lists the types of the total number of instruments currently used by the partner hospital and the percentage of each type. Note that about 24% of the instruments used are loaned or consigned.

Table 1.26: A List of Instruments Used

Type	**Total Number**	**Total in %**
Loaner	266	5%
Consigned	1,095	19%
Owned	3,507	61%
Other Services	927	16%
Total	5,795	100%

This study evaluates three approaches to the delivery of surgical supplies to ORs and compares their performance. These approaches were designed through discussions with the GMH staff. Three groups of performance measures are used to compare these approaches: (*i*) the average delay in the start time of surgery and the corresponding frequency, which measure the level of service provided by the OR, (*ii*) the number of instruments inventoried, which measures the efficiency of the inventory system, (*iii*) the total daily travel time, the average travel time per trip and the corresponding standard deviation, which measures the efficiency of material handling systems.

Table 1.25: AGV Movements by the Time of the Day

Route	Time Interval	Average No. of Daily Trips	Travel Time [Min] Average	Travel Time [Min] Std. Dev.	Coefficient of Variation
Materials Department - Case Cart Storage Area	12 am-3 am	2.6	4.66	10.26	2.2
	3 am-6 am	4.5	5.96	9.55	1.6
	6 am-9 am	2.2	5.27	6.17	1.17
	9 am-12 pm	1.6	5.8	2.4	0.41
	12 pm-3 pm	2.0	5.33	3.69	0.69
	3 pm-7 pm	**28.3**	**8.94**	**6.49**	0.73
	7 pm-9 pm	5.1	5.67	4.45	0.78
	9 pm-12 am	3.9	4.88	5.72	1.17
2nd Floor Soiled Cart Storage - CSSD	12 am-3 am	0.9	7.28	6.16	0.85
	3 am-6 am	1.1	6.87	6.16	0.9
	6 am-9 am	2.9	5.55	3.27	0.59
	9 am-12 pm	19.6	4.56	4.04	0.88
	12 pm-3 pm	17.6	5.17	2.56	0.49
	3 pm-7 pm	15.1	**9.71**	**8.51**	0.88
	7 pm-9 pm	2.5	6.65	3.04	0.46
	9 pm-12 am	1.5	5.45	1.01	0.19

The first approach, Model 1, is called the *Current* approach and assumes that the materials required by the surgeries are delivered to the CCSA the night before surgery. The *Current* approach is the ongoing practice of the partner hospital in the data presented here.

Next, the *Two Batch* approach, Model 2, assumes that the materials required for the surgeries scheduled in the morning are delivered to the CCSA the previous evening, and the materials required for the surgeries scheduled in the afternoon are delivered in the morning on the day of surgery. This approach provides the opportunity to reuse instruments from the surgeries scheduled later in the day. Since the CSSD works 24 hours each day, the instruments can be washed overnight and delivered in the morning. Since instruments are delivered a few hours in advance of surgery, staff have a long window of time to intervene if an instrument becomes unavailable. Therefore, the risk of not getting the instruments on time is only minimal and does not affect the quality of care in the hospital.

Finally, the *Just-in-Time* approach, Model 3, assumes that the materials required are delivered shortly before the start of the surgery. The time between the delivery of surgical supplies and the surgery, referred to as the delivery interval, must be determined and affects inventory levels. The required inventory level increases with the delivery interval. For example, consider two surgeries that require the same instrument and are scheduled on the same day. In the current system, a hospital must have two sets of identical instruments, since they are delivered to the CCSA the day before surgery. If the delivery interval is 1 hour, the instrument can be sterilized and delivered before the subsequent surgery, provided that the two surgeries are scheduled several hours apart. In this case, the hospital needs only one instrument. However, if the delivery interval is chosen to be less than 3 hours, then, to avoid any delays in the second surgery, two instruments are needed. Note that the implementation of JIT and other lean methods in healthcare, unlike manufacturing, should be considered with caution because such practices could delay surgeries and jeopardize the well-being of patients.

The delays of a surgery start time are separated into two categories: delays due to AGVs, e.g., long travel time due to congestion or unavailability of AGVs, and delays due to unavailable instruments. Delays due to AGVs can create challenges for the JIT approach, but these delays can be reduced by optimizing the fleet size. Delays due to unavailable instruments are caused by delays in the delivery of the soiled case carts or by an increase in the number of emergency surgeries. A delay in the delivery of the soiled case carts subsequently delays the cleaning process of the instruments and carts, which delays the start of the next surgery that uses the same instrument. Delays due to unavailable instruments can be reduced by optimizing the inventory level. In this research, simulation experiments are conducted to determine the optimal fleet size and inventory level under each proposed material handling approach. On the basis of the results of these experiments, the delivery interval that optimizes the identified performance measures is also determined. In order to reduce the computational time of simulation-optimization experiments, a lower bound, based on the data collected at GMH, is developed on the number of instruments inventoried. Let n be the maximum number of surgeries scheduled in a day for each type of service. The lower bound equals $\lceil n/2 \rceil$. A lower bound is added for each type of surgery through these constraints: (*i*) the number of instruments used $\leq$ number of instruments in the inventory and (*ii*) the number of instruments used $\geq$ lower bound. We note that the composition of surgical trays affects the level of inventory of surgical instruments. However, this research does not focus on tray composition.

Each of the proposed material handling approaches requires a different number of AGVs to deliver materials on time. This number is affected by the surgical schedule and the material handling process. For example, the number of AGVs needed for the JIT delivery approach is lowest, since the delivery of the case carts is spread throughout the day. The number of AGVs needed by the *Current* delivery approach is larger because the delivery of case carts is completed in a short period of time. The number of AGVs needed also depends on the total number of cases scheduled and the spread of the schedule. A tight schedule would require more AGVs to complete material handling on time.

Limitations of this research:

Model: The research proposed here is conducted in collaboration with GMH, a US-based hospital located in Greenville, South Carolina. The models presented here are motivated by the material handling and inventory management practices of GMH. The research team worked closely with the perioperative services department, which consists of the MD, the CSSD, and the OR Department. GMH uses AGVs to transport surgical case carts to and from ORs. The problem setting proposed here and the assumptions made are influenced by the practices at GMH. The models presented here are a valuable contribution to the literature because, based on a careful review of the literature, other hospitals follow similar practices for material handling and inventory management of surgical instruments.

Data: Nine months of real-life data are used to develop the case study. These data include information about the number of surgical cases each day and span a time period long enough to observe how seasonality impacts the number of surgical cases. Ideally, larger amounts of data would be available, but that does not apply here.

1.5.1 Simulation Model

DES models are developed to evaluate and compare the three approaches proposed for the delivery of surgical supplies. These models are created in ARENA simulation software by Rockwell Automation. An entity type represents a surgical type, and each entity represents a surgical case of a particular type. An entity has three attributes: *duration, starting time, and type*. *Duration* is randomly generated using distributions derived from the data collected at GMH. The *starting time* and *type* are fed to the model from the actual data. Other entities are used to control the movement of AGVs and elevators, as well as to handle other specific requirements, such as calculating the value of certain variables (e.g., the number of AGVs to activate each day). ORs, case carts, cart washers, and elevators were modeled as resources. Variables are used to track the number of resources used. We used the same guided path transporter network developed with intersections and links to replicate the movement of AGVs along the hospital corridors. This network was constructed using actual distances obtained from a GMH floor map. The links in the network are unidirectional, bidirectional, or spurs (dead ends). The intersections represent the areas where two or more links intersect. The intersections allow the AGVs to turn and move from one link to the next, following their routes. Intersections are also used to represent pick-up/drop-off stations. A spur link marks the end of a route. Departments can only handle a certain number of AGVs, and their processing capacity is limited by variables.

The first DES model, Model 1, shows the material handling approach *Current* used in GMH. Figure 1.14 describes this model. In this model, the release of entities begins at 3 pm. The start time of these entities takes place after 6 am the next day. Next, the availability of instruments is checked using the decide module. If an instrument is not available, the case cart is held in the MD until the instrument becomes available. An available instrument seizes a case cart and is delivered to the CCSA. There, the entity is held until the scheduled start time of the surgery. At this point in time, the entity seizes an available OR for the duration of the surgery. At the end of a surgery, the OR is released and the corresponding case cart and instrument are moved to the CSSD to wash and sterilize. Resources in CSSD are seized for the duration of service. The variables that record the number of busy units are updated when resources are released. The second DES model, Model 2, depicts the *Two-Batch* material handling approach. In this model, the entities are released twice a day, at 6 am and 3 pm. Entities released at 6 am have a start time between after 12 pm the same day. Next, these entities follow a similar procedure as described above in Model 1. Entities released at 3 pm have a start time between 6 am and noon the next day. These entities are held until the next morning using the hold module, and then they follow the procedure described above. The third DES model, Model 3, shows the *JIT* approach. In this model, entities are released one hour before their *start time*. This delivery interval was chosen on the basis of the results of our numerical analysis. Next, these entities follow the same procedure as described above. In all models, the delivery of soiled carts begins as soon as surgery is completed.

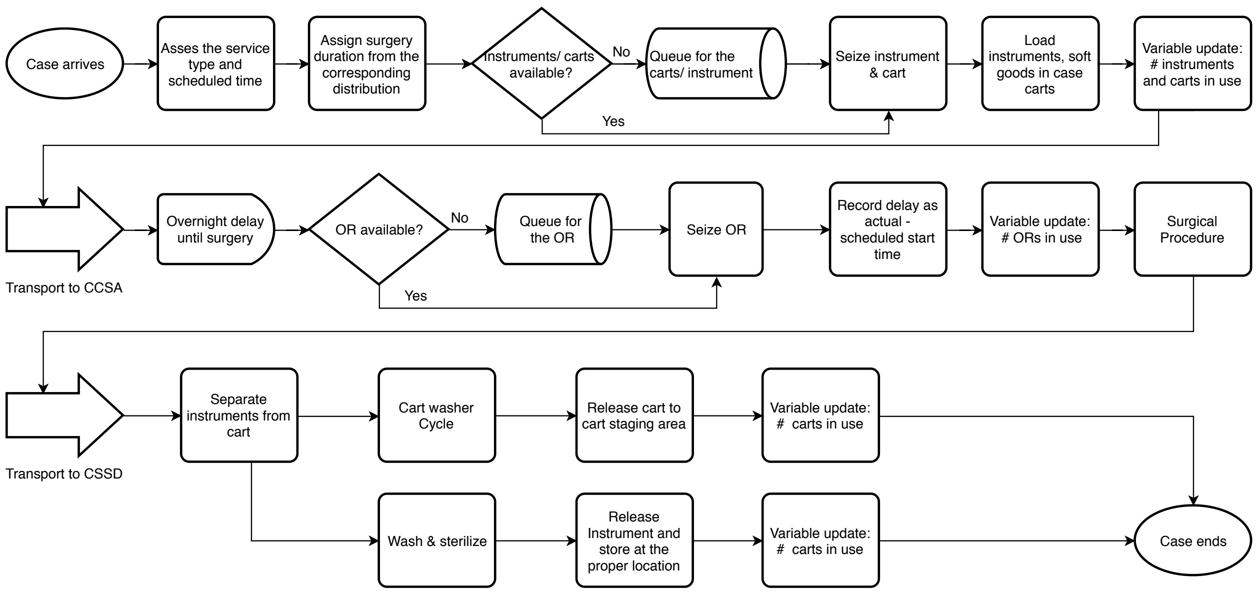


Figure 1.14: Flowchart of the Simulation Model

1.5.2 A Case Study

In this subsection, we describe a case study based on the data obtained from GMH. This study evaluates the impact of the three proposed approaches on material handling system, surgical delays, surgical instrument utilization, and surgical delays. Based on the results of our data analysis, we generate the parameters required for the-implementation of our case study via simulation. This data allowed for a thorough statistical analysis of the results. In our first set of experiments, the inventory level was varied to three levels, that is, low, medium, and high. The three models are compared to each other in terms of surgical delays due to material handling. Next, by fixing a realistic surgical delay amount, which is seen in the current system, experiments were carried out to determine the level of inventory required for other approaches. These approaches are also tested for material handling system performance indicators such as travel times, surgical delays due to the material handling system, and frequency of delays.

Input Data Analysis: The main objective of data collection and analysis is to support simulation models. These models then evaluate the impacts that the *Current* material handling approach has on the surgical instrument inventory. Information collected, such as the OR schedule, instrument inventory levels, and instrument requirements, is used directly as input to the simulation model. Our partner hospital provided two sets of data. The first data set provides detailed information on the surgeries scheduled from January 1, to September 11, 2018. These data include the surgical identification number (ID), OR ID, the date of surgery, the scheduled start and end times of a surgery, the type of surgery (i.e., vascular, orthopedic, neurological, etc.), information about the surgeon, the primary procedure, and the instruments requested. The second data set provides information about the surgical instruments used. These data present the instrument ID, the type of surgery the instrument is used for, the level of inventory, and information about its ownership. Tables 1.27 to 1.30 summarize these data. The hospital offers 46 different surgical services. Our experimental study focuses on the following seven types of surgery: ENT, pediatric, ortho trauma, neurology, gynecology, urology, and vascular. We focus on these surgical services because they are scheduled multiple times a day, and therefore there is an opportunity to reduce the size and cost of inventory by reusing some of the instruments. Surgeries are grouped based on service type, duration, and scheduled start times. The duration of a surgery is calculated using the actual start and finish times. For each service type, an hypothesis test is conducted to assess whether the duration of surgeries within each type of service differs based on the start time of the given surgery. When differences were observed, the distribution of the duration of surgery was estimated separately. Otherwise, the data was used to derive a single distribution for surgeries of the same type that were started at different times of the day. The results of the hypothesis test generated the input parameters used in the simulation model. For example, the duration of surgery differs based on the time of day the surgery is scheduled, by day of the week, and also by type of service. A continuous distribution was fitted using the Input Analyzer of Rockwell Automation to represent the duration of surgery. Table 1.27 shows the service types, distribution of the length of surgeries, and the squared error. The real-life scheduled start times of the surgeries are used in the simulation model obtained from the data set and presented here.

Table 1.28 summarizes the total number of surgical cases scheduled between January 1 and September 11, 2018. Here, only the surgery types that were scheduled more frequently are listed. Each of these types of surgery is scheduled more than once a day and requires multiple instruments of the same kind. For each type of surgery, only one set of instruments is considered, common to all surgical cases of that type. Table 1.29 lists the instruments selected for this study and their corresponding inventory levels. Other data used to develop the simulation model are summarized in Table 1.30.

GMH carries multiple instruments for each type of surgery for three main reasons: First, the same surgery could be scheduled more than once on the same day if the hospital follows a block-schedule approach. This approach assigns

Table 1.27: Input Parameters: Surgery Duration

Service	From	To	Distribution	Expression (Length of Surgery)	Squared Error
ENT Surgery	00:00	08:00	Lognormal	LOGN(2.02, 2.12)	0.008
	08:00	14:00	Lognormal	LOGN(1.62, 1.2)	0.007
	14:00	00:00	Lognormal	LOGN(1.23, 0.672)	0.003
Gynecology Service	07:00	08:00	Beta	0.01 + 4.81 * BETA(2.85, 4.03)	0.009
	15:00	16:00	Lognormal	0.27 + LOGN(0.965, 0.511)	0.005
	16:00	07:00	Lognormal	LOGN(1.65, 0.859)	0.011
Neurological Surgery	00:00	09:00	Gamma	GAMM(0.494, 5.44)	0.007
	09:00	13:00	Erlang	ERLA(0.454, 5)	0.005
	13:00	00:00	Beta	12 * BETA(4.95, 25.8)	0.028
Ortho Trauma Surgery	0:00	8:00	Erlang	ERLA(0.587, 5)	0.002
	08:00	14:00	Lognormal	LOGN(2.57, 1.29)	0.004
	14:00	00:00	Lognormal	LOGN(2.09, 0.983)	0.004
Pediatric Surgery	00:00	00:00	Lognormal	LOGN(1.35, 0.693)	0.011
Urology Surgery	00:00	07:00	Lognormal	LOGN(1.63, 0.975)	0.001
	07:00	08:00	Lognormal	LOGN(1.2, 0.821)	0.007
	08:00	00:00	Erlang	ERLA(0.244, 4)	0.011
Vascular Surgery	00:00	07:00	Beta	0.03 + 8.97 * BETA(0.97, 1.78)	0.004
	07:00	09:00	Gamma	GAMM(0.608, 3.58)	0.017
	09:00	14:00	Gamma	GAMM(0.42, 4.39)	0.025
	14:00	00:00	Triangular	TRIA(0.13, 0.83, 3.54)	0.011

Table 1.28: Input Parameters: Number of Surgeries

Service	**Sunday**	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Total**
ENT Surgery	25	295	148	264	231	302	20	1,285
Gynecology Service	17	227	133	181	176	198	13	945
Neurological Surgery	22	174	168	293	163	225	17	1,062
Ortho Trauma Surgery	2	205	171	174	206	207	56	1,021
Pediatric Surgery	62	145	248	153	276	158	79	1,121
Urology Surgery	39	293	333	298	382	466	72	1,883
Vascular Surgery	61	141	242	224	241	235	81	1,225
Total	228	1,480	1,443	1,587	1,675	1,791	338	8,542

the same block of time every week to a surgeon or a group of surgeons who perform the same type of surgery. These surgeons use instruments of same service type. The current material handling system requires that all instruments be available one day before surgery. Second, surgeons of different specialties may request the same instrument for the same procedure. Third, the hospital carries safety stock to respond to instrument-related incidents, such as dropping or breaking an instrument during surgical procedures.

Table 1.29: Number of Instruments in the Inventory

Service Type	Instrument	Inventory
ENT Surgery	Set T & A GMMC 1047	10
Gynecology Service	Set D & C mini GMMC 15896	10
Neurological Surgery	Set Back Neuro GMMC 1341	12
Ortho Trauma Surgery	Set Minor Ortho GMMC 100031	17
Pediatric Surgery	Set Pediatric Minor GMMC 1247	8
Urology Surgery	Ureteroscope 7.5 Comp GMMC 12656	18
Vascular Surgery	Probe Doppler Pencil 8.1 GMMC 1824	25

Table 1.30: Summary of Input Parameters

Parameter	Source	Description
Entity Creation Time	Surgery schedule data	Read from the data
Attribute Duration	Surgery schedule data	Random variable from the corresponding distribution
Network link distances	GMH floor maps	Read from the data
No. of Case carts	AGV system data	110
No. of ORs	GMH Survey	32
No. of loading personnel	GMH Survey	4
No. of AGVs	AGV system data	[6,8,10]
Capacity of elevators	GMH Survey	[2,2]
Capacity of cart washers	GMH Survey	3
Cart loading delay	GMH Survey	Triangular(2,3,5) minutes
Cart washing delay	GMH Survey	20 Minutes
Elevator movement delay to carry AGV	GMH Survey	40 seconds
Cart loading unloading delay	GMH Survey	15 seconds
Instrument washing delay	GMH Survey	3 hours

Verification and Validation: Verification and validation procedures are used to compare the conceptual model with the proposed DES models. The development of the DES models is guided by the process flow chart and uses input data provided by GMH staff, who examined and approved these models. Because we validated the simulation of the

AGV material handling system previously, it is not required to be revalidated. Furthermore, the approach proposed by [132] is adopted to verify and validate the DES models. The input data analysis section describes our data collection and analysis. This analysis indicates that our data are correct and used adequately. The conceptual model proposed is validated via *face validation* by GMH staff and via *traces* following specific entities through the model. Flowcharts of the conceptual model are verified by GMH staff. The DES models are verified through the techniques listed in [132]. These techniques include animation, comparison with other models, and running several replications of the model. In the sensitivity analysis, the number of resources used (i.e., the number of AGVs, the number of instruments, etc.) changed, so the impact of these changes on the behavior of the model outputs was monitored. The model outputs considered are the average travel time of an AGV and the corresponding standard deviation. Next, hypothesis tests were conducted to evaluate whether the difference between the output of the DES models and the real-world data is statistically different. Using a significance level of 0.05, the test indicates that the difference is not statistically significant.

1.5.3 Discussion of Results

In this subsection, we present the results obtained from the experiments carried out as part of the case study. The objective of this case study is to evaluate the impact of the three proposed approaches on the material handling system, surgical delays, the utilization of surgical instruments, and surgical delays. We discuss how these experiments address the research questions outlined previously.

Research Question 3: How does the inventory level of surgical instruments, including owned, borrowed, and cosigned, impact the service level provided by the ORs? A simulation-optimization experiment is conducted using ARENA Opt-Quest to answer this question. The objective of simulation-optimization is to minimize the total delays at the start of a surgery by changing the inventory level. The delay of a surgery is calculated as the difference between the *Actual Start Time* and the *Scheduled Start Time*. The decision variables of type *integer* are the number of instruments in the inventory for each of the seven types of service (see Table 1.29). Experiments are conducted for three different scenarios. Scenario 1 assumes that the available inventory of instruments is equal to the current inventory level of GMH. Consider this inventory level to be an upper bound. Scenario 2 assumes that the available inventory of instruments is equal to the lower bound. Scenario 3 assumes that the available inventory of instruments equals the average value of the upper and lower bounds. To ensure that the inventory level is an integer, we round the average value down to the nearest integer. Table 1.31 summarizes the results of these experiments, and the following observations result:

Observation 1: The *Current* material handling approach, Model 1, is the most sensitive to changes in inventory level, compared to *Two Batch*, Model 2, and *JIT*, Model 3. A decrease of inventory level, from

Scenario 1 to Scenario 2, leads to an increase in the average delay from 0.42 to 31 minutes per surgery in the *Current* approach. In the *Two Batch* approach, the corresponding average delay increases from 0.01 to 5.12 minutes, and in the *JIT* approach from 0.00 to 1.47 minutes per surgery (see Table 1.31).

Observation 2: The *Current* material handling approach requires additional levels of inventory to maintain the same level of service, measured by the average delay per surgery, compared to the proposed *Two Batch* and *JIT approaches* (see Table 1.32).

Observation 3: The *JIT* approach leads to reduced inventory levels of instruments used in short-duration surgeries without reducing the service level.

Table 1.28 presents the total number of neurological surgeries performed in GMH during the 9-month period reviewed here. This number averages about 4.2 surgeries per day. Table 1.28 also presents the number of pediatric surgeries during the same time period, which corresponds to approximately 4.4 surgeries per day. The duration of neurological surgeries is about 1 hour longer than for pediatric surgeries. A hypothesis testing (significance level = 0.05) was conducted to evaluate the difference between the duration of neurological and pediatric surgeries. This test indicates that the difference is statistically significant (see Table 1.33). The results of Table 1.32 show that the number of instruments required by neurological surgeries is higher than pediatric surgeries in Models 2 and 3 versus Model 1. This is because instruments used in pediatric surgeries can be reused on the same day due to the shorter duration of these surgeries.

Table 1.31: The Average Delay per Surgery

	Number of Instruments per Service Type							**Average Delay/Surgery (Minutes)**		
Scenario	**ENT**	**Gynecology**	**Neurological**	**Ortho Trauma**	**Pediatric**	**Urology**	**Vascular**	**Model 1**	**Model 2**	**Model 3**
1	10	10	13	17	8	18	16	0.42	0.01	0
2	6	6	5	9	4	10	6	31.27	5.12	1.47
3	8	8	9	13	6	14	10	3.58	0.25	0.01

Table 1.32: Inventory Level of Instruments

		Number of Instruments per Service Type						
Model	**Average Delay/ Surgery (Minutes)**	**ENT**	**Gynecology**	**Neurological**	**Ortho Trauma**	**Pediatric**	**Urology**	**Vascular**
1	0.42	10	10	13	17	8	18	16
2	0.41	6	10	13	13	6	18	12
3	0.41	6	6	9	9	4	16	12

Table 1.33: Comparison of Two Service Types

Statistics	Neurology Surgery	Pediatric Surgery
Sample Size	1062	1121
Average Length	2.31	1.36
95% CI	(2.24,2.38)	(1.31,1.40)
Standard Deviation	1.16	0.77

Research Question 4: How do material handling activities impact the service level provided by the ORs? Two sets of experiments are carried out. The first set focuses on the impact that changing the number of AGVs has on the performance of the material handling system measured via the average travel time per trip, the total travel time and the corresponding standard deviations. The delivery time of clean and soiled case carts is analyzed as the number of AGVs increases from 6 to 8 to 10. Experiments with fewer than six AGVs caused extensive delays in delivering all case carts in the *Current* system, which requires employees to work overtime, so these experiments are not considered in this analysis. The second set of experiments focuses on the impact that changing the delivery time has on the performance of the material handling system. For this purpose, the performances of Models 1, 2, and 3 are compared. The results of these experiments are summarized in Tables 1.34 and 1.35 and Figures 1.15 and 1.16. The following observations result:

Observation 1: The average daily travel time of clean case carts is longest in the *Current* material handling approach and shortest in the *JIT* approach (see Figures 1.15a, 1.15b, and 1.15c). The *Current* approach has the longest travel time due to congestion, as the delivery of clean case carts for elective surgeries takes place between 3-6 p.m.

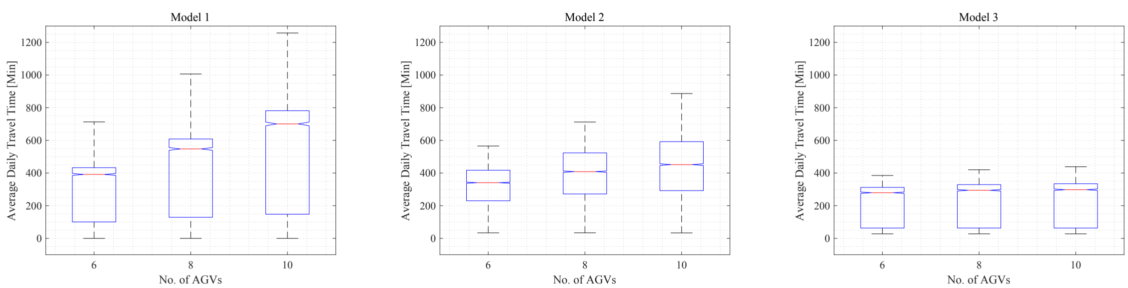
Observation 2: The average daily travel time of clean case carts increases with the number of AGVs (see Figures 1.15a, 1.15b, and 1.15c). This increase is highest in the *Current* material handling approach.

Observation 3: The average daily travel time of clean case carts in the *JIT* approach is not affected by the increase in the number of AGVs since the delivery of case carts is spread throughout the day. These deliveries do not cause congestion (see Figure 1.15c).

Observation 4: The average daily travel time of soiled case carts for every material handling approach is slightly affected by the increase in the number of AGVs (see Figures 1.16a, 1.16b, and 1.16c). Note that the difference in average travel time per trip is small but still statistically significant. The change in travel time due to the increase in the number of AGVs for every Model is small because soiled case carts are delivered to CSSD right after the surgery; thus, they are delivered throughout the day, and these deliveries have a minimal impact on congestion.

Table 1.34: Sensitivity Analysis of Clean Case Carts Delivery

		Travel Time (Minutes)			
No. of AGVs	**Model**	**Average**	**StDev**	**CI for Average**	**CI for StDev**
6	**1**	4.83	0.31	(4.83, 4.84)	(0.30, 0.31)
	2	4.96	0.17	(4.95, 4.96)	(0.17, 0.18)
	3	3.31	0.27	(3.31, 3.32)	(0.26, 0.27)
8	**1**	6.53	0.60	(6.51, 6.54)	(0.59, 0.61)
	2	6.47	0.64	(6.46, 6.48)	(0.62, 0.65)
	3	3.43	0.36	(3.42, 3.44)	(0.35, 0.36)
10	**1**	8.02	1.16	(7.99, 8.05)	(1.14, 1.17)
	2	7.66	1.17	(7.63, 7.69)	(1.14, 1.19)
	3	3.46	0.39	(3.46, 3.47)	(0.38, 0.39)



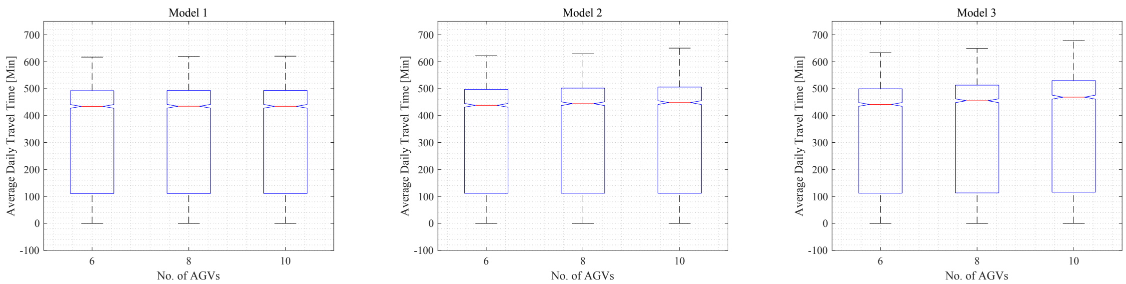
(a) Total Travel Time: Model 1 (b) Total Travel Time: Model 2 (c) Total Travel Time: Model 3

Figure 1.15: Sensitivity Analysis of Clean Case Carts Delivery

Table 1.35: Sensitivity Analysis of Soiled Case Carts Delivery

		Travel Time (Minutes)			
No. of AGVs	**Model**	**Average**	**StDev**	**CI for Average**	**CI for StDev**
6	**1**	5.46	0.10	(5.46, 5.46)	(0.094, 0.097)
	2	5.44	0.09	(5.44, 5.44)	(0.087, 0.089)
	3	5.39	0.07	(5.39, 5.39)	(0.069, 0.071)
8	**1**	5.60	0.18	(5.59, 5.60)	(0.179, 0.186)
	2	5.53	0.15	(5.53, 5.54)	(0.149, 0.154)
	3	5.40	0.07	(5.39, 5.40)	(0.072, 0.074)
10	**1**	5.75	0.29	(5.75, 5.76)	(0.290, 0.299)
	2	5.64	0.23	(5.63, 5.64)	(0.232, 0.238)
	3	5.40	0.07	(5.39, 5.40)	(0.072, 0.074)

Research Question 5: How does the integration of decisions about inventory and material handling impact the service level provided by the ORs?



(a) Total Travel Time: Model 1 (b) Total Travel Time: Model 2 (c) Total Travel Time: Model 3

Figure 1.16: Sensitivity Analysis of Soiled Case Carts Delivery

To address this research question, we compared the three models with each other in terms of surgical delays and the level of inventory required to meet the requirements. We show how the amount of surgical delays and their frequency depend on the number of AGVs and the delivery approach chosen. The following observation results:

Observation: The average delay per surgery and the total number of delays are lowest in the *JIT* material handling approach. These statistics are highest in the *Current* approach. A successful implementation of *JIT* requires coordination of material handling and inventory decisions, and numerical results show that this coordination leads to improved service level provided by ORs.

This observation is true at different inventory levels, represented by Scenarios 1, 2, and 3 in Table 1.31; for different material handling approaches, represented by Models 1, 2, and 3 in Tables 1.36 and 1.37; and for different material handling capacities, represented by the number of AGVs in Tables 1.36 and 1.37.

Observation 2: The average inventory level is lowest in the *JIT* material handling approach. This observation is supported by the results of Tables 1.31 and 1.32.

Table 1.36: Average Delay by Service Type (Hours)

	Model 1			Model 2			Model 3		
Service Type	**6 AGV**	**8 AGV**	**10 AGV**	**6 AGV**	**8 AGV**	**10 AGV**	**6 AGV**	**8 AGV**	**10 AGV**
ENT	0.060	0.006	0.001	0.009	0.000	0.000	0.001	0.000	0.000
Gynecology	0.132	0.011	0.001	0.014	0.000	0.000	0.000	0.000	0.000
Neurological	1.672	0.124	0.005	0.330	0.011	0.000	0.085	0.000	0.000
Ortho trauma	0.018	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
Pediatric	1.186	0.244	0.042	0.266	0.020	0.002	0.030	0.001	0.000
Urology	0.245	0.024	0.001	0.014	0.000	0.000	0.000	0.000	0.000
Vascular	0.324	0.003	0.000	0.028	0.000	0.000	0.003	0.000	0.000

*The total number of replications is 30.

Table 1.37: Frequency of Delayed Surgeries

	Model 1			Model 2			Model 3		
Service Type	**6 AGV**	**8 AGV**	**10 AGV**	**6 AGV**	**8 AGV**	**10 AGV**	**6 AGV**	**8 AGV**	**10 AGV**
ENT	378	41	5	175	14	2	17	0	0
Gynecology	580	60	5	138	10	0	10	0	0
Neurological	6,254	595	22	3,352	314	0	1,820	0	0
Ortho Trauma	120	0	0	63	0	0	0	0	0
Pediatric	4,859	1,293	255	2,532	377	52	743	20	0
Urology	2,261	268	8	652	39	0	8	0	0
Vascular	1,722	24	0	615	5	0	85	0	0

*The total number of replications is 30.

Recommendations: The following recommendations are made based on the observations presented above: Coordinating material handling and inventory management decisions has the potential to improve the level of service provided by ORs. To facilitate this coordination, the requirements for each surgery, the number of available instruments, and the location of the instruments must be known at all times. Transparent information technology systems will facilitate the coordination of decisions. Hospitals should consider implementing a *JIT* material handling approach for instruments used in short-duration surgeries because such an approach leads to lower inventory levels without jeopardizing the level of service provided. Finally, hospitals should frequently reevaluate their material handling system to identify improvements. For example, GMH currently uses 10 AGVs. Using only 6 or 8 AGVs leads to reduced congestion along hospital corridors and leads to shorter delivery times. The remaining AGVs can be used to transport trash, linen, and pharmaceuticals, among other items.

1.6 Conclusions

The inefficiencies observed in the GMH material handling system motivated this research. GMH staff reported long lines of AGVs waiting for the elevator on the mezzanine floor after material handling activities started in the afternoon. Congestion caused by AGVs contributes to delays in delivering the required surgical material, including surgical instruments to ORs. Congestion also affects the delivery of soiled surgical instruments, which further delays sterilization activities. Delays due to material handling activities impact the utilization of AGVs, surgical instruments, sterilization equipment, and personnel time. These delays also force GMH to use the rental and consigned instruments.

Our careful review of the literature indicates that several other hospitals also follow similar practices for material handling and inventory management of surgical instruments. New data-based approaches to material handling and inventory management, such as the delivery of surgical cases *JIT*, have the potential to reduce surgical delays due to instrument unavailability in the ORs. Furthermore, the utilization of surgical instruments can be improved for short-duration surgeries in hospital ORs. Hospitals should identify opportunities to coordinate material handling and inventory management decisions, as it leads to reduced delays due to the material handling system, improved utilization of inventoried surgical instruments, and reduced usage of rental and consigned instruments.

Congestion due to material handling activities can be addressed by improving the material handling infrastructure or improving the schedule of material handling activities. Data analysis indicates that the number of trips, the average travel time per trip, and the corresponding standard deviation of travel times are higher in the part of the day when clean surgical carts are delivered. Furthermore, congestion was observed on the AGV paths shared for different tasks. A simulation-optimization model with alternative paths for AGVs was developed. The use of alternative routes led to reduced congestion. However, the overall travel time increased due to longer travel times. Finally, additional investments to install AGV guide-paths on the new routes would be required to implement this redesign. A sensitivity analysis was performed to evaluate the impact of reducing the number of AGVs on the average travel time and the task completion time. This analysis indicates that the use of fewer AGVs is sufficient to complete the daily delivery of surgical carts. For example, using more than 6 AGVs does not significantly improve task completion time. However, the travel time per trip increases with the number of AGVs due to increased congestion. Furthermore, the number of surgical cases affects the optimal number of AGVs that should be used on one day and the next. This suggests that using 6 AGVs for each day of the week may not be optimal. To account for these daily differences, we conducted two simulation-optimization experiments and identified the number of AGVs to be used. The results suggest the use of a different number of AGVs on different days of the week. To evaluate the results of the simulation-optimization model, a solution was selected that used fewer AGVs every day and yields a task completion time that met the delivery target. This selected solution was simulated for a longer period of time using historical data as input. The number of AGVs used per day was fixed on the basis of the results of the proposed framework. Statistical analysis of the results indicated that the proposed model led to reduced congestion and shorter travel times.

In addition to numerical experimentation using our simulation models, we conducted a one-week pilot study at our partner hospital. During this study, the number of AGVs used in GMH was the same as that in the selected solution. The travel times for the treatment week were compared with those for the previous week and the week after, during which the number of AGVs in the system was not controlled. This comparison indicates that travel times were shorter during the week-long pilot study. The extensive data analysis and results of the simulations presented here reinforce what GMH staff already suspected: increasing the number of AGVs can exacerbate material handling issues rather than alleviating them. These results indicate that the number of AGVs should change daily according to the volume of surgical cases.

At our partner hospital, the material handling system currently delivers case carts loaded with instruments the evening before the surgery. This restrictive delivery schedule is one of the reasons the hospital cannot reuse surgical instruments on the same day. This delivery schedule leads to (i) increased inventory requirements for instruments; (ii) increased traffic and congestion from AGVs, which delay case cart deliveries and other materials that use AGVs; and (iii) delayed surgery start times. To address these operational problems, we proposed two new material handling approaches and compared them with current practice. The *Two Batch* approach delivers surgical carts to ORs twice a day, requiring partial coordination of material handling and inventory management. The *JIT* approach takes a surgical cart to an OR before surgery that requires complete coordination of material handling and inventory management. The simulation models were verified and validated using real-life data collected from the partner hospital. A thorough sensitivity analysis leads to a number of observations and recommendations. The *Current* material handling approach is more sensitive to changes in inventory level, requires the highest levels of inventory to reduce instrument-related surgical delays, and leads to congestion and delays in the delivery of surgical case carts. Both the *Two Batch* and *JIT* approaches outperform the current material handling approach. Implementing the *JIT* approach leads to the greatest improvements in terms of surgical delays, AGV utilization, and surgical instrument utilization.

The staff at our partner hospital have considered the models and experimental results as valuable inputs, and have implemented our recommendations in some capacity. These recommendations also played an important role in helping hospital management assess possible changes to be made in the design of the facility.

Chapter 2

Models and Solution Methods for Integrated OR Scheduling and Inventory Management Problem

2.1 Introduction

In 2013, healthcare expenditures exceeded $2.9 trillion, which accounted for 17.4% of the gross domestic product (GDP) of the USA that year. In 2017, these expenditures increased to $3.5 trillion and made up 17.9% of the USA's GDP [114]. Studies project that national healthcare expenditures will increase by 60% during the period 2005-2025 [112] and might climb at an average annual rate of 5.4% during 2019-2028, whereas prices for medical services and goods are expected to increase at an average annual rate of 2.4% in the same period [86]. The trend of increasing healthcare expenditures negatively impacts the US economy and Americans' well-being [12].

US National Health Expenditure Accounts show that in 2017, 62% of expenditures were due to hospital care, physicians, clinical services, and prescription drugs [114]. In 2018, total spending for hospital care increased 4.5 %, which amounted to $1.3 trillion [86]. Moreover, it is expected that during 2024–28, the total hospital spending will continue to rise by 6% per year. Surgical operations account for a significant proportion of the costs and revenues of a hospital. In 2011, 29% of the total 38.6 million hospital stays involved surgical procedures, and these stays accounted for nearly 50% of the total $387 billion spent on hospital costs in the USA [158]. Surgical expenditures are expected to increase from $572 billion in 2005 (4.6% of USA GDP) to $912 billion in 2025 (7.3% of US GDP) [112]. Moreover, a 2014 study conducted in acute care hospitals in California suggests that the average total cost of using an operating room (OR) varies between $36 and $37 per minute [26], indicating that operating room costs are a considerable part of overall surgical expenditures.

There is a vast amount of research that focuses on reducing the cost of healthcare by improving the efficiency of operations in hospitals. Some of this research focuses on improving the efficiency of OR, staff scheduling, inventory management, material handling, etc. Our research contributes to this literature by developing models that coordinate decisions about OR scheduling and inventory management. We demonstrate that our proposed model has the potential to reduce the cost of ORs, and consequently reduce the cost of healthcare.

Background: The fixed cost of operating an OR is significant because of the cost of OR staff and the supporting upstream personnel. Typically, ORs have planned session lengths of eight to ten hours per day. Using an OR beyond the session length results in *overtime* costs, while not using an OR results in *idle-time* costs. Overtime and idle-time also result in staff dissatisfaction. Between two surgeries, cleaning and setup activities are performed to prepare for the next surgery. The time spent on these activities is called *OR turnover time* [16]. Each surgeon requires a set of surgical instruments to perform surgeries. The instrument requirement is specified by a doctor's preference card (DPC). After each surgery, instruments are sterilized. The time spent on these activities is referred to as *instrument turnover time*. There is a cost associated with the transportation and sterilization of surgical instruments. We refer to this cost as *usage cost*. The OR-related costs are significantly higher than instrument-related costs. Therefore, hospitals have prioritized improving the utilization of ORs. For example, previous studies find that personnel cost is $12-$14 per minute [26]. In comparison, usage cost is found to be less than $3.5 per instrument [43, 148]. The difference in costs explains why there have not been many studies that evaluate the impacts of OR scheduling on the cost of instrument inventory.

Improper instrument inventory management decisions have their own repercussions. For example, if the required surgical instruments are not available before the surgery begins, the hospital is forced to postpone the surgery. These delays can be minimized by purchasing additional instruments or using rental instruments. While carrying excess inventory seems to be a good option to maintain high service levels, the costs associated with a large surgical instrument inventory are hefty. The inventory holding cost of surgical instruments can account for 4% to 8% of the hospital's operating budgets [41, 156]. To avoid holding excess inventory, hospitals often depend on vendors to supply specialized instruments as *rental* instruments. Rental instruments remain the property of the vendor until the instrument is sold. The hospital pays vendors a premium for using and maintaining rental instruments [99]. The vendors prefer renting instruments as they get to use the hospital's shelf space to store their instruments while generating revenue upon use.

Hospitals must pay close attention to how the OR scheduling decisions influence the inventory-related decisions. For example, a study by [99] finds that orthopedic and cardiovascular supply vendors had overfilled storage shelves in a major US hospital with items that were rarely being used. Furthermore, the rental instruments were overstocked by more than 50% as compared to owned instruments, which increased the total cost by 12-25%. A similar situation was observed at our industry partner, Greenville Memorial Hospital (GMH), located in South Carolina. The data on the instrument usage at GMH between *January-September, 2018* was collected and analyzed. Figure 2.1 shows the total number of instruments rented at GMH. Figure 2.2 presents the number of surgical instruments rented per week by different services. These figures reveal that GMH has used a large number of rental instruments for Neurological and Orthopedic surgeries every week. The instruments used for these services are the most expensive [22]. GMH has used at least 50 rental instruments per week for both services. Our research work is motivated by this overuse of rental instruments at GMH.

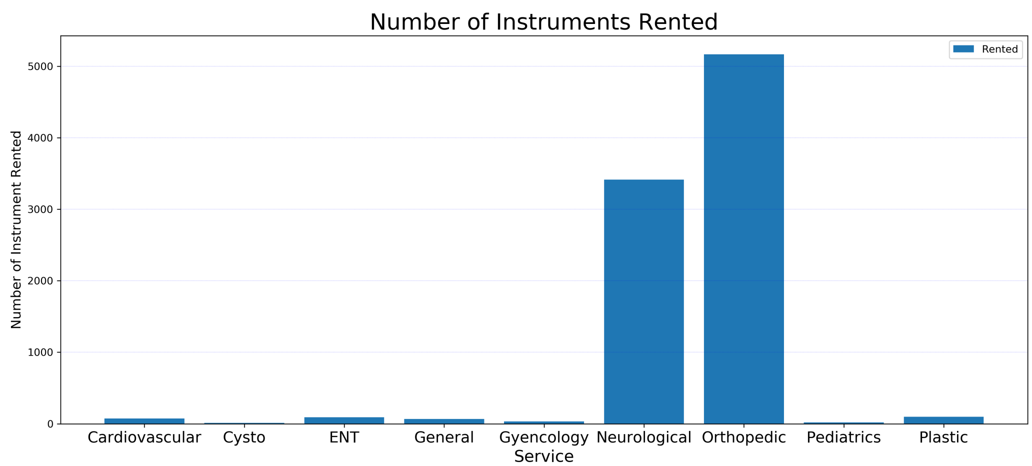


Figure 2.1: Number of Instruments Rented

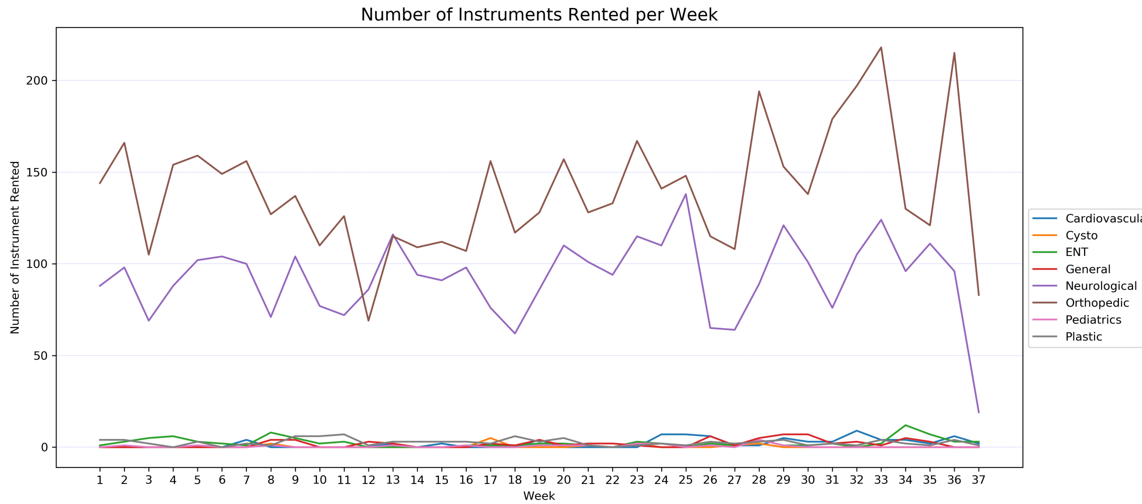


Figure 2.2: Number of Instruments Rented per Week

Research Questions: Currently, the hospital generates an initial surgery schedule in advance, and next, the hospital determines what instruments are available and what should be rented. We call this the *advanced scheduling* approach. There are a number of caveats that concern the cost-effectiveness of these *advanced* OR schedules. An important consideration in designing the OR schedule is inventory level and availability of instruments. In a multi-OR environment, surgeries are performed in parallel, therefore, the surgical instruments are used concurrently. If the available instrument inventory level cannot meet the requirements, rental instruments are used and the hospital incurs high costs. The availability of instruments is also impacted because of the sterilization process. Instruments must be sterilized in the sterilization department (SD) after every use. This means that instruments are unavailable for a few hours after each use before they can be used again. Accidental dropping of instruments in ORs, failures of sterilization processes, missing instruments, and delivering incorrect instruments into ORs are some of the other reasons that can adversely affect the advance OR schedules. In such situations, hospitals resort to postponing surgeries or using rental instruments incurring a higher cost. These challenges motivate our first research question: *(i) How does integrating decisions about OR scheduling and inventory management impact the total costs of the system?* We propose a mathematical model to integrate the decisions of OR scheduling with the assignment of instruments to surgeries. With effective planning, this coordination can increase the utilization of the ORs and the surgical instruments leading to cost savings for hospitals. In the long run, the inventory level is impacted by the total number of surgeries scheduled in a day, the daily schedule of surgeries that use the same instrument, the processing capacity of the sterilization department (SD), and the schedule of material handling activities. Changing demand patterns, surgeons' requests for additional/newer equipment, or damage to existing instruments influence inventory levels. To minimize the use of rental instruments, the OR manager must consider the inventory levels in determining the sequence of surgeries scheduled in an OR, and the number of surgeries scheduled in parallel. Since the surgical instruments are cleaned overnight, it seems intuitive that scheduling surgeries over multiple days provides hospitals more flexibility in using the inventory. However, this may not be entirely possible due to some practical constraints. In some cases, OR managers must respect the patient-surgeon assignments, patient preferences, and surgeons' preferences for scheduling the surgery. These problems motivate our second research question *(ii)* How does the inventory level of surgical instruments, the length of planning horizon, and number of ORs scheduled in parallel impact the system utilization and the cost of an integrated OR schedule? We conduct a sensitivity analysis to evaluate how the parameters mentioned above impact the utilization of ORs, and instruments, and impact the cost of the system.

Illustrative Examples: To illustrate how our proposed integrated OR scheduling works, we provide two examples. Consider that a hospital plans to schedule four surgeries. The duration and instrument requirements for each surgery are provided in Table 2.1. Figure 2.3 summarizes the potential schedules of each example. In this Figure, each block represents a surgery, and the size of a block depicts the duration of the surgery. The subscript denotes the copy of the instrument to be used. The first schedule proposed for Example 1 (see Figure 2.3(a)) is an *advanced* schedule of ORs. In this schedule, surgeries 1, 2, and 3 are assigned to OR 1, while surgery 4 is assigned to OR 2. Schedule 1 in Figure 2.3(a) presents a feasible instrument assignment for the *advanced* schedule, where instruments B_1 and A_1 are assigned to surgeries 2 and 4 respectively. This assignment forces the hospital to use rental instruments of type A for surgeries 1 and 3. This is because surgeries 1 and 4 are scheduled at the same time. Since it takes 3 hours to sterilize an instrument, instrument A_1 cannot be reused during surgery 3. In contrast, Schedule 2 uses instrument A_1 for surgeries 1 and 3. Thus, to implement this schedule, the hospital must rent 1 instrument. Example 1 illustrates that even if OR schedules are generated in advance, hospitals can save on the rental costs by strategically allocating the instruments.

ID	Duration	Instrument Type Needed
1	2	A
2	2	A
3	4	B
4	4	B

Table 2.1: Data for Illustrative Examples

Example 2, depicted in Figure 2.3(b), uses the same data provided in Table 2.1. In this example, we assume that additional copies of surgical instruments are available. Schedule 3 uses instruments A_2 and B_2 for surgeries 2 and 4.

This is because instruments A_1 and B_1 can not be reused since there is not enough time to sterilize them. In Schedule 4, instrument A_1 is used for surgery 1, it is sterilized while surgery 3 is in progress, and it is reused in surgery 2. These schedules demonstrate that by *sequencing* the surgeries differently, hospitals can increase the utilization of surgical instruments, and lower inventory levels. Schedules 4 and 5 can be implemented using only one copy of instruments A and B. This is because, when used in different days (i.e., instruments used for surgeries 1 - A_1 and 3 - B_1), instruments are sterilized overnight. These schedules illustrates that by scheduling surgeries on separate days, the hospital has more flexibility in using the instruments. The second copies of instruments can now be used to schedule add-on surgeries. These examples show that coordinating OR scheduling with instrument assignments can help hospitals efficiently use their surgical instruments and meet their instrument requirements with fewer rented instruments.

Surgery	1		2				3		OR 1
Surgery	4								OR 2
Hour	1	2	3	4	5	6	7	8	

Advanced Schedule

Surgery	1 - R		2 - B_1				3 - R		OR 1
Surgery	4 - A_1								OR 2
Hour	1	2	3	4	5	6	7	8	

Schedule 1

Surgery	1 - A_1		2 - B_1				3 - A_1		OR 1
Surgery	4 - R								OR 2
Hour	1	2	3	4	5	6	7	8	

Schedule 2

(a) Example 1

Surgery	1 - A_1		2 - A_2						Day1
Surgery	3 - B_1				4 - B_2				Day2
Hour	1	2	3	4	5	6	7	8	

Schedule 3

Surgery	1 - A_1		3 - B_1				2 - A_1		Day1
Surgery			4 - B_1						Day2
Hour	1	2	3	4	5	6	7	8	

Schedule 4

Surgery	1 - A_1		3 - B_1						Day1
Surgery	2 - A_1		4 - B_1						Day2
Hour	1	2	3	4	5	6	7	8	

Schedule 5

(b) Example 2

Figure 2.3: Illustrative Examples

Contributions: The proposed research offers several important contributions: (*i*) This study showcases the role of coordinating OR scheduling and inventory management decisions in improving OR efficiency and reducing costs. In particular, this work demonstrates that integrating these decisions can increase the utilization of ORs and surgical instruments while reducing the cost of the system. Prior works point to the cost savings and benefits of optimizing

OR scheduling process. However, based on our review of the literature, only a few papers discuss the inclusion of instrument inventory-related decisions in OR schedules. (*ii*) This work develops solution approaches to solve this problem efficiently. We propose easy-to-implement solution methods including a construction heuristic and a Lagrangean decomposition-based heuristic and evaluate them. These solution approaches outperform the commercial solver, Gurobi, in terms of running time and solution quality. (*iii*) This work develops a real-life case study using data from a US-based hospital. The proposed models and methods, which are intuitive and easy to implement, are used to derive managerial insights that can lower the cost of health care. While the models presented here are particularly suitable for the hospitals that use an open scheduling strategy, other healthcare facilities can also learn from these practices.

The rest of this chapter is organized as follows: Section 2.2 reviews the literature that is relevant to this work. Section 2.3 provides a detailed description of the problem and its formulation. Section 2.4 describes the proposed solution methods. Section 2.5 introduces a case study and discusses the results of the computational experiments. Finally, section 2.6 summarizes the key takeaways and presents concluding remarks.

2.2 Literature Review

The focus of this research is to generate elective surgery schedules and inventory assignments in a multi-day multi-OR setting. The main streams of literature relevant to this research are inventory management of reusable surgical instruments and OR scheduling. In this section, we provide a brief review of the literature, identify the important knowledge gaps, and contrast related studies with our work.

2.2.1 Inventory Management of Reusable Surgical Instruments

The literature related to inventory management of surgical instruments focuses on minimizing the costs associated with their purchase, use, and inventory. Most of the inventory management strategies proposed by practitioners are practical and easy to implement. These strategies include improving the accuracy of doctor preference cards (DPC), minimizing the number of unused instruments in surgical trays, and getting surgeons involved in the cost reduction process. For example, [74] engaged physicians in the review of DPCs, which led to the removal of 109 disposable supplies and 3 reusable instrument trays. Consequently, the cost of a case cart was reduced by $16 on average. These findings align with the research by [81], which stipulates that surgeons often underestimate the costs of expensive items and overestimate the costs of less expensive items. Research by [62, 14] shows that increasing surgeon awareness about the costs of surgical instruments and equipment can lower the cost of surgical procedures. Work by [138] shows that standardizing surgical processes can also significantly reduce operating costs. These easy-to-implement approaches that focus on involving surgeons in the cost reduction process, reduced the cost of surgical procedures. We note that these studies assume that OR schedule is known. However, the daily instrument requirements are affected by these schedules. Our research extends the scope of these models by developing models that coordinate surgery scheduling decisions with inventory management. For a further systematical review of the literature in the area of inventory management of surgical instruments, the readers are referred to [4].

Several researchers have studied the optimal composition of surgical trays to reduce the cost of inventory. This problem is referred to as the tray optimization problem (TOP). The TOP consists of three important decisions: *(i)* the assignment of instruments to trays, *(ii)* the assignment of trays for surgeries, and *(iii)* the number of trays to keep in inventory [42]. Surgeon preferences for instruments are also taken into account. Works by [41, 127, 5, 147, 42] have presented exact solution approaches to solve TOP, and works by [41, 42] have used heuristic/metaheuristic methods. Although minimizing the cost of inventory is the main focus of TOP, researchers have extended the classical TOP model to consider objectives such as minimizing delivery costs, storage costs [147, 127], processing costs [127], wastage [127, 5], and the number of trays [5]. Our work does not focus on the configuration of surgical trays. We note that our approach extends the scope of instrument/tray assignments to surgeries by coordinating these decisions with OR scheduling.

2.2.2 OR Scheduling

The OR scheduling and planning approaches are classified into two categories in the literature, namely advance scheduling and allocation scheduling [166]. Advance scheduling determines well in advance the exact date when a surgery is scheduled, and/or the OR where the surgery takes place. For example, [71] propose an advance OR scheduling model that determines the assignment of surgeries to ORs over a one to two weeks planning horizon. They solve the resulting problem using a primal-dual-based heuristic. [39] study the deterministic and stochastic versions of the multi-OR scheduling problem to minimize the cost of ORs. Their formulation does not determine in what sequence the surgeries are scheduled in each OR. They propose easy-to-implement heuristics to solve this problem. Allocation scheduling is mainly used to determine the starting time of each surgery and to allocate the OR resources. For example, [101] analyze the impact of OR sequencing rules on the utilization of the recovery unit. The sequencing rules are applied independently to each surgeon's list of cases. They show via a discrete event simulation model that, sequencing rules of OR have a great impact on the quality of care that patients receive in the recovery room. [94] present a scheduling strategy to determine starting times of surgeries in multiple ORs. They use a genetic algorithm to determine the allocation of surgeries to ORs and the starting time of each surgery. A few research articles propose models that integrate advance scheduling and allocation scheduling decisions. For example, [102] propose a model that schedules each surgery into a day of the week, into an OR, and into a time slot in order to maximize OR efficiency. Works by [146] and [54] present a multi-phase approach to solve the OR planning and scheduling problem. The initial phases focus on the advanced scheduling problem, and then the sequence of surgeries is determined. [16] consider decisions such as the number of ORs to open, surgery assignment to ORS, and the starting time of each surgery in their daily OR scheduling problem. Their model highlights the importance of OR pooling and parallel processing of surgeries. [15] propose a deterministic mathematical model that yields daily OR schedules to minimize the OR-related costs as well as costs of recovery resources. A decomposition-based heuristic and an easy-to-implement two-phase heuristic are proposed to solve this problem. Similar to their work, we formulate an integrated advance and allocation scheduling problem. However, our proposed model, in addition to integrating advance scheduling and allocation scheduling decisions, considers several decisions related to the inventory of surgical instruments.

There are three different strategies used to schedule ORs, i.e., the block strategy, the open strategy, and the modified block strategy [166]. The block scheduling strategy divides OR time in blocks and allocates each block to a surgeon or to a surgical group in advance. This strategy reduces the complexity of scheduling by reducing the problem size because each surgery can only be scheduled within the corresponding block of time. The block scheduling strategy is used by many hospitals in Europe [121] and in the US [149, 152, 107, 68, 53]. However, a key issue with block scheduling is the low utilization of certain blocks since a surgeon cannot schedule surgeries in blocks that are pre-allocated to other surgeons even if they are empty [166]. In the modified block strategy, a block can be modified if under-utilized. Research by [38, 54] presents models that can be used to determine OR schedules using a modified block strategy. The open scheduling strategy provides surgeons the flexibility to schedule a surgical case on any day in any available OR. This strategy also allows surgeries of different specialties to be scheduled in the same OR. However, a poorly designed open scheduling strategy could cause issues and inconveniences for the hospital during an emergency. Such a strategy could also inconvenience surgeons when their surgeries are scheduled at different times of the day [166]. [13] use an open scheduling strategy where patients are scheduled without specialty constraints. They propose a Lagrangean relaxation-based method to solve this problem. Similarly, [52, 54, 75] have adopted an open scheduling strategy in their work. Our proposed scheduling model follows an open scheduling strategy due to the flexibility it provides in scheduling surgeries so that the inventory costs are minimized.

Some authors include general instrument-related constraints in their formulation of the OR scheduling problem. For example, [19, 21, 152] have included constraints either ensure availability of surgical instruments or instrument turnover (sterilization) time constraints. [19] extend the OR scheduling problem to consider surgical teams, instruments, and recovery resources. They solve this problem in two stages *(i)* OR assignment, *(ii)* allocation of the resources. Their results indicate increased OR occupancy, increased number of surgeries performed daily, and improved utilization of ORs and recovery resources. They consider a similar problem setting to ours and assume that instruments are sterilized immediately after surgery. However, unlike our model, their formulation does not capitalize on the benefits of scheduling over multiple days. Work by [21] studies a multi-objective combinatorial optimization problem to determine the sequence of patients within the OR. Their constraints ensure that number of instruments in use does not exceed the inventory levels. They also include instrument turnover time in their model. Work by [91] formulates the OR scheduling problem to include the recovery resources, and equipment/instruments for both elective and emergency surgeries. They formulate an integer linear programming (ILP) model to solve small instances of their problem, and then use constraint programming and metaheuristics to solve larger instances. They limit the number of surgeries scheduled in parallel that require the same instruments. Their model does not consider instrument turnover time, unlike our formulation. [152] address multi-period, multi-resource, priority-based OR scheduling. They propose a mixed-integer programming (MIP) model and present a heuristic based on the first fit decreasing algorithm. The objective of their model is to maximize the number of patients scheduled. Inventory level and availability constraints are included in the model. They report that this proposed approach led to substantial savings and increased utilization at a publicly funded hospital. Work by [107] proposes a constraint programming model that includes several real-life constraints, such as availability of instruments/equipment, staff preferences, and affinities among staff members. The objectives are to minimize the make-span and overtime hours, and to maximize affinities among the team members. They ensure the availability of surgical instruments in the schedule they develop. We note that the general instrument related constraints considered in the articles we cite here are also critical for our model, and included in our formulation. However, the aforementioned articles do not focus on the opportunity to reduce the cost of healthcare operations by coordinating surgical instrument assignments and OR scheduling. We address this gap in the literature with the purpose of reducing the cost of surgical instrument inventory.

There are only a handful of other researchers who have placed an importance on the integration of OR scheduling and instrument assignment decisions to reduce costs. Most of the literature focuses in reducing the cost of ORs because these costs are higher than the cost of surgical instruments. However, careful OR scheduling can minimize the costs associated with surgical instruments [6, 30]. The authors in [6] solve a monthly integrated multi-OR scheduling problem. They propose an MIP model and a robust formulation with the objective of minimizing overtime, the number of ORs opened, and the number of instruments sterilized in an emergency. There are some key differences between their and our work. Firstly, they assign the surgeons to ORs. Each surgeon is assigned to a fixed list of surgeries to perform in the same OR. Therefore, their model uses a block schedule. Secondly, the sterilization schedules and related constraints are tailored to the business practices of their partner hospital. Finally, the problem is solved in a lexicographic fashion. Their results show that integrated OR scheduling significantly reduces OR costs and also minimizes emergencies at their partner hospital. Work by [30] integrates decisions about scheduling of multi-ORs with the decisions about sterilization of surgical instruments. They propose an MIP model to minimize the costs of sterilization, postponement of surgeries, and the makespan. They show that developing OR schedules first is not beneficial when the schedule is affected by inventory decisions. They also report that a large percentage of problem instances result in infeasibility when sterilization decisions were included after the OR schedule was generated. The authors propose a batch-based heuristic that decomposes the problem into two stages. The first stage assigns surgeries to ORs and batches, and then the second stage sequences the surgery via an iterative procedure. It is reported that the integrated OR scheduling reduced the total OR costs. However, the authors do not capitalize on the benefits of scheduling surgeries over multiple days. Different from [30], our proposed model allows us to exploit the trade-offs between using OR overtime, purchasing and using inventoried instruments, and using rental instruments.

2.3 Problem Description and Formulation

Consider a hospital that seeks to determine the weekly schedule of ORs. Assume that the surgeries have deterministic durations and known instrument requirements. We also assume that the hospital carries several instruments of each type in the inventory. The requirement for surgical instruments can either be met via inventoried instruments or via rental instruments borrowed on a per surgery basis. We develop an MIP model to identify a schedule of ORs by assigning each patient to an OR at a specific time and day within the given time horizon. In addition to determining the OR schedule, our model decides whether to rent an instrument or not, the number of instruments to rent, and an assignment of rental/inventoried instruments to each surgery. The OR schedule and the instrument assignments identified by the proposed model minimize the total cost of the system. The cost parameters we include in our model are designed to match the reality for most ORs in hospitals in the USA. These parameters include the cost of opening an OR, the cost associated with OR overtime and idle-time, the cost of using an instrument from the inventory, and instrument rental costs. Overtime represents the number of time slots an OR was occupied beyond the determined session length. Similarly, idle-time occurs when an OR is opened, but not occupied.

Hospitals strategically invest in standardized, flexible OR suites to promote operational efficiency. Therefore, our model allows ORs to be used for more than one specialty. After a surgery, a turnover is required to clean and set up the OR for the next scheduled surgery. We include the OR turnover time in surgery duration. The sterilization time for surgical instruments is accounted for separately in our model. We assume that surgical instruments are made available right before the surgery and material handling delays are ignored. Finally, we assume that cancellations are not allowed. We choose a time index-based formulation i.e. we break up time into discrete time slots $\mathbb{T} = 1 \dots T$. The parameters such as duration of surgeries, sterilization time for instruments, etc. are given in terms of the number of time slots. Time index-based formulation is helpful in tracking the whereabouts of instruments and patients as part of the solution methods we propose.

Notations:

Sets and Indices

$k \in \mathbb{K}$	Set of ORs
$i \in \mathbb{N}$	Set of surgeries
$t \in \mathbb{T}$	Set of time slots
$d \in \mathbb{D}$	Set of days
$m \in \mathbb{M}$	Set of instruments
$r \in \mathbb{R}$	Set of types of instruments

Parameters

l_i	Duration of surgery i in time slots
s	Session length of OR ($s < T$)
c^u	Cost of idle-time \$/time slot
c^o	Cost of overtime \$/time slot
c^k	Cost of opening an OR \$/use
c_r^F	Cost of using an instrument of type r \$/use
c_r^R	Cost of borrowing an instrument of type r \$/use
γ	Number of time slots required for instrument turnover
d_{ir}	Number of instrument of type r required by surgery i
p_{mr}	$\begin{cases} 1, & \text{if instrument } m \text{ is of type } r \\ 0, & \text{otherwise} \end{cases}$
I_r	Inventory level of instruments of type r

Decision Variables

$$X_{idtk} \quad \begin{cases} 1, & \text{if surgery } i \text{ is assigned to OR } k \text{ on day } d \text{ and starts in time slot } t \\ 0, & \text{otherwise} \end{cases}$$

$$Z_{idtk} \quad \begin{cases} 1, & \text{if time slot } t \text{ on day } d \text{ in OR } k \text{ is occupied by surgery } i \\ 0, & \text{otherwise} \end{cases}$$

$$Q_{dk} \quad \begin{cases} 1, & \text{if OR } k \text{ is opened on day } d \\ 0, & \text{otherwise} \end{cases}$$

$$V_{im} \quad \begin{cases} 1, & \text{if surgery } i \text{ uses instrument } m \\ 0, & \text{otherwise} \end{cases}$$

$$W_{imtd} \quad \begin{cases} 1, & \text{if instrument } m \text{ is unavailable due to being used by surgery } i \text{ in time slot } t \text{ on day } d \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha_{imtd} \quad \begin{cases} 1, & \text{if surgery } i \text{ starts in time slot } t \text{ on day } d \text{ and uses instrument } m \\ 0, & \text{otherwise} \end{cases}$$

O_{dk} the amount of overtime on day d in OR k

U_{dk} the amount of idle-time on day d in OR k

R_{ir} the number of instruments of type r rented for use in surgery i

The following is an MIP formulation of the integrated OR scheduling problem (P):

(P):

$$\min Z_p = \sum_{d=1}^{\delta}\sum_{k=1}^{K}\left[c^o O_{dk} + c^u U_{dk} + c^k Q_{dk}\right] + \sum_{i=1}^{N}\sum_{r=1}^{R}\left[c_r^R R_{ir} + \sum_{m=1}^{M} c_r^F\, p_{mr} v_{i,m}\right] \quad (2.1a)$$

Subject to :

$$\sum_{i=1}^{N} X_{idtk} \leq Q_{dk}, \qquad \forall d,t,k, \quad (2.1b)$$

$$Q_{dk} \geq Q_{dk'}, \qquad \forall k,k' \in \mathbb{K}, k' > k, d, \quad (2.1c)$$

$$\sum_{d=1}^{\delta}\sum_{i=1}^{N}\sum_{k=1}^{K} Z_{idtk} = \sum_{d=1}^{\delta}\sum_{k=1}^{K} Q_{dk}, \qquad \forall t, \quad (2.1d)$$

$$\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} X_{idtk} = 1 \qquad \forall i, \quad (2.1e)$$

$$\sum_{d=1}^{\delta}\sum_{t=T-l_i+1}^{T}\sum_{k=1}^{K} X_{idtk} = 0 \qquad \forall i, \quad (2.1f)$$

$$t * Z_{idtk} \leq s + O_{dk} \qquad \forall d,k,t, \quad (2.1g)$$

$$s * Q_{dk} - \sum_{i=1}^{N}\sum_{t=1}^{T} l_i * X_{idtk} \leq U_{dk} \qquad \forall d,k, \quad (2.1h)$$

$$\sum_{t'=t}^{t+l_i-1} Z_{idt'k} \geq l_i * X_{idtk} \qquad \forall i,d,k, 1 \leq t \leq T - l_i + 1, \quad (2.1i)$$

$$\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} Z_{idtk} = l_i \qquad \forall i, \quad (2.1j)$$

$$\sum_{i=1}^{N} Z_{idtk} \leq 1 \qquad \forall t,d,k, \quad (2.1k)$$

$$R_{ir} + \sum_{m=1}^{M} p_{mr} V_{im} = d_{ir} \qquad \forall i,r, \quad (2.1l)$$

$$\sum_{i=1}^{N} W_{imtd} \leq 1 \qquad \forall t \in 1 \dots T+\gamma, m, d, \quad (2.1m)$$

$$\sum_{i=1}^{N}\sum_{m=1}^{M} p_{mr} W_{imtd} \leq I_r \qquad \forall t,d,r, \quad (2.1n)$$

$$\sum_{t=1}^{T} W_{imtd} = (l_i + \gamma) * \sum_{t=1}^{T} \sum_{k=1}^{K} X_{idtk} * V_{im} \qquad \forall i, m, d \tag{2.1o}$$

$$\sum_{t'=t}^{t+l_i+\gamma-1} W_{imt'd} \geq (l_i + \gamma) * X_{idtk} * V_{im} \qquad \forall i, k, d, 1 \leq t \leq T - l_i - \gamma + 1, \tag{2.1p}$$

$$X_{idtk} \in \{0,1\}, \qquad Z_{idtk} \in \{0,1\}, \qquad Q_{dk} \in \{0,1\}, \qquad O_{dk} \geq 0, \qquad U_{dk} \geq 0, \tag{2.1q}$$

$$W_{imtd} \in \{0,1\}, \qquad R_{ir} \in Z^{+}, \qquad V_{im} \in \{0,1\}. \tag{2.1r}$$

The objective function (2.1a) minimizes the total cost of scheduling the ORs, which consists of, the cost of overtime, idle-time, opening an OR, using instruments, and renting instruments. Constraints (2.1b) ensure that a surgery is assigned to an OR only if the OR is open. Constraints (2.1d) ensure that the total number of surgeries does not exceed the number of ORs open at any time. Since all ORs have the same planned session length, there is complete symmetry with respect to ORs. Thus, for any solution, an equivalent solution can be obtained by swapping the sets of surgeries assigned to any pair of ORs on the same day. Constraints (2.1c) eliminate the symmetry and force ORs to be opened in order. These constraints reduce solution space, thus, making the problem easier to solve [15, 135, 39]. Constraints (2.1e) and (2.1f) make sure that each surgery is scheduled, and is completed before the end of the day. Constraints (2.1g) calculate OR overtime, which occurs when the OR is occupied beyond the planned session of length S. Constraints (2.1h) calculate the unused OR time. Constraints (2.1i) ensure that an OR is considered busy during the surgery. That is, if surgery i begins at period t of day d in OR k, then, no other surgery will be scheduled in OR k during periods t to $t + l_i$ of day d since it is in use by surgery i. Constraints (2.1j) ensure that the number of periods a surgery is scheduled for equals the duration of the surgery. Constraints (2.1k) ensure that at most one surgery can be scheduled during a particular time slot in a given OR. Constraints (2.1l) ensure that the requirement for a resource type is met either via rental or inventoried instruments. Constraints (2.1m) ensure that an instrument is used in only one surgery at any time slot. Constraints (2.1n) ensure that in any time slot, the number of inventoried instruments being used does not exceed the inventory levels. Constraints (2.1o) and (2.1p) ensure that if an instrument is used for surgery, then it is busy (i.e., unavailable to be used for another surgery) during the surgery and the instrument turnover time.

The formulation given above is a nonlinear MIP due to the bilinear term $X_{idtk} * V_{im}$ in constraints (2.1o) and (2.1p). We linearize these bilinear terms using the McCormick relaxation method [105]. This results in an exact reformulation since both X_{idtk}, and V_{im} are binary variables. We introduce the binary decision variable α_{imtd} which takes value 1 if surgery i begins in time slot t of the day d and uses instrument m, and takes value 0 otherwise. To ensure that α_{imtd} takes value 1 when both $X_{idtk} = 1$ and $V_{im} = 1$, we add the following set of constraints to the model.

$$\alpha_{imtd} \geq 0 \qquad \forall i,t,m,d, \qquad (2.2a)$$

$$\alpha_{imtd} \geq \sum_{k=1}^{K} X_{idtk} + V_{im} - 1 \qquad \forall i,d,t,m, \qquad (2.2b)$$

$$\sum_{t'=t}^{t+l_i+\gamma-1} W_{imt'd} \geq (l_i+\gamma) * \alpha_{imtd} \qquad \forall i,d,m, 1 \leq t \leq T - l_i - \gamma + 1, \qquad (2.2c)$$

$$\sum_{t=1}^{T} W_{imtd} = (l_i+\gamma) * \sum_{t=1}^{T} \alpha_{imtd} \qquad \forall i,m,d. \qquad (2.2d)$$

This provides the final formulation of problem (P).

2.4 Solution Methods

Problem (P) is a challenging MIP model to solve directly. In this section, we develop alternative solution methods including a construction heuristic (H1) and a Lagrangean decomposition-based heuristic (LDH).

2.4.1 Construction Heuristic (H1)

Heuristic H1 solves the problem in two phases. In phase 1, we use an MIP formulation for the *OR scheduling* problem. We refer to this formulation as model (S). The solution to model (S) determines the assignment of surgeries to days and ORs. The start times of surgeries are also determined by ensuring that surgeries adhere to the scheduling constraints. In phase 2, we fix this OR schedule in model (P). The resulting model with a fixed schedule is an MIP formulation for the *instrument assignment* problem. We refer to this formulation as model (F).

The model (S) for the *OR scheduling* problem is given below. This model uses the same notations as model (P).

(S):

$$Z_S = \min \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^o O_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^u U_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^k Q_{dk} \qquad (2.3)$$

$$\text{Subject to: } (2.1b)-(2.1k), (2.1q)$$

The objective function (2.3) minimizes the fixed cost of opening the ORs and the variable costs of OR overtime and idle time.

Solving model (S) in the first phase of heuristic H1 generates an OR schedule. To enforce this schedule in model (F), the values of OR scheduling variables $X, Z, Q, O,$ and U from model (S) are fixed as $\hat{X}, \hat{Z}, \hat{Q}, \hat{O},$ and $\hat{U}$ respectively in model (P). As a result, OR scheduling constraints (2.1b) - (2.1k) are not a part of model (F). We include the remaining constraints of model (P) in model (F).

The model (F) for the *instrument assignment* problem is given below. This model uses the same notations as model (P).

(F):

$$Z_F = \min \sum_{d=1}^{\delta}\sum_{k=1}^{K}\left[c^o\hat{O}_{dk} + c^u\hat{U}_{dk} + c^k\hat{Q}_{dk}\right] + \sum_{i=1}^{N}\sum_{r=1}^{R}\left[c_r^R R_{ir} + \sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{d=1}^{\delta} c_r^F p_{mr}\alpha_{imtd}\right] \tag{2.4}$$

$$\text{Subject to : } (2.1l)-(2.1n), (2.2a), (2.2c), (2.2d), (2.1r)$$

A solution to model (F) i.e. (α, W, V, R) provides an optimal surgical instrument assignment for the OR schedule obtained from model (S). The proposition we present next shows that this is a feasible assignment for model (P).

Proposition 1. *If the tuple (X, Z, Q, O, U) is a feasible solution to model (S), and $(alpha, W, V, R)$ is a feasible solution to model (F), then $(X, Z, Q, O, U, \alpha, W, V, R)$ is a feasible solution to model (P).*

Proof. Given a feasible solution to (S), we can always construct a solution to (F). This is because there is no upper limit on the number of rental instruments that can be used. Therefore, an instrument assignment from owned inventory can always be augmented with rental instruments to meet the instrument requirements. Since (F) uses a feasible schedule and generates a feasible instrument assignment, it provides a feasible solution to (P). □

H1 uses models (S) and (F) to generate a feasible solution to model (P). The main drawback of this method is that the quality of the solution is highly dependent on the quality of the OR schedule obtained from model (S).

2.4.2 Lagrangean Decomposition-Based Heuristic

The Lagrangean decomposition (LD) approach developed by [70] has been used to solve optimization models for supply chain planning and scheduling problems [48, 145, 163, 111, 20], production scheduling problems [161, 69, 33], transportation planning [48, 49] etc. Since this method is effective for solving large-scale MIP problems, we propose an LD-based heuristic (LDH) to generate high-quality solutions for (P).

LDH works as follows. First, we reformulate (P) to facilitate the use the LD approach. For this reformulation, we introduce new variables, which are simply *copies* of certain variables in (P). We also include *copy constraints* to enforce the equality of these copy variables to the original ones. We refer to the extended model that results from the reformulation of (P) as model (Q). The inclusion of copy constraints and copy variable makes model (Q) larger, however, the optimal solution of model (Q) is the same as that of model (P). Next, we dualize the copy constraints that will lead to decomposition into subproblems (SP1) and (SP2). Each subproblem is solved separately. Since we have a minimization problem, the sum of the objective function values of the subproblems will be lower than the optimal objective function value of (P), and can therefore be used as a lower bound (LB) for (P). As the next step, we apply a heuristic that uses the solutions from (SP1) to generate feasible solutions to (P) that can be used as an upper bound (UB) for the optimal solution of (P). Once we obtain an upper bound, the subgradient method is used to update the values of the Lagrangean multipliers. The subproblems are again solved using the new multiplier values to obtain new LB and UB. This process is repeated until certain stopping criteria are met. The main idea of this decomposition method is to separate the integrated OR scheduling problem into two subproblems that are computationally easier to solve.

2.4.2.1 Problem Reformulation

In this subsection, the reformulation of (P) is described. We present new variables and constraints required for the reformulation. First, we introduce new binary variables Y_{idtk}, that are *copies* of the scheduling variables X_{idtk} in (P). Constraints (2.5c) enforce the equality of these two sets of variables. Using these variables, we rewrite a set of constraints in (P) by replacing X_{idtk} by Y_{idtk} as shown in (2.5b). This leads to the following equivalent extended formulation of (P) i.e. model (Q).

(Q):

$$\begin{aligned} \min \ & \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^{o} O_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^{u} U_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^{k} Q_{dk} + \sum_{i=1}^{N}\sum_{r=1}^{R} c_r^R R_{ir} \\ & + \sum_{r=1}^{R}\sum_{d=1}^{\delta}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} c_r^F p_{mr} \alpha_{imtd} \end{aligned} \tag{2.5a}$$

Subject to: equations (2.1b) - (2.1n), (2.2a), (2.2d), (2.1q), and (2.1r),

$$\alpha_{imtd} \geq \sum_{k=1}^{K} Y_{idtk} + V_{im} - 1 \qquad \forall i \in \mathbb{N}, t \in \mathbb{T}, m \in \mathbb{M}, d \in \mathbb{D}, \tag{2.5b}$$

$$X_{idtk} = Y_{idtk} \qquad \forall i \in \mathbb{N}, d \in \mathbb{D}, t \in \mathbb{T}, k \in \mathbb{K}, \tag{2.5c}$$

2.4.2.2 Lagrangean Decomposition

As the first step of our decomposition scheme, we dualize the copy constraints (2.5c) and take these constraints to the objective function. The following is the Lagrangean decomposition (LD) formulation of model (P):

$$\min \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^o O_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^u U_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^k Q_{dk} + \sum_{i=1}^{N}\sum_{r=1}^{R} c_r^R R_{ir} + \sum_{r=1}^{R}\sum_{d=1}^{\delta}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} c_r^F p_{mr}\alpha_{imtd} - \sum_{i=1}^{N}\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} \lambda_{idtk}(Y_{idtk} - X_{idtk}) \tag{2.6}$$

Subject to: equations (2.1b) - (2.1n), (2.2a), (2.2c), (2.2d) and (2.5b).

The values of the multipliers λ_{idtk} can be positive, negative, or zero, due to equality constraints. The formulation (LD) can now be separated into the following two subproblems:

(SP1):

$$Z_{LD1}(\lambda) = \min \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^o O_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^u U_{dk} + \sum_{d=1}^{\delta}\sum_{k=1}^{K} c^k Q_{dk} + \sum_{i=1}^{N}\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} \lambda_{idtk} X_{idtk} \tag{2.7}$$

Subject to: equations (2.1b) - (2.1k), and (2.1q).

(SP2):

$$Z_{LD2}(\lambda) = \min \sum_{i=1}^{N}\sum_{r=1}^{R} c_r^R R_{ir} + \sum_{r=1}^{R}\sum_{d=1}^{\delta}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} c_r^F p_{mr}\alpha_{imtd} - \sum_{i=1}^{N}\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} \lambda_{idtk} Y_{idtk} \tag{2.8a}$$

$$\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K} Y_{idtk} = 1 \qquad \forall i, \tag{2.8b}$$

$$\sum_{d=1}^{\delta}\sum_{t=T-l_i+1}^{T}\sum_{k=1}^{K} Y_{idtk} = 0 \qquad \forall i, \tag{2.8c}$$

Subject to: equations (2.1l) - (2.1n), (2.2a), (2.2c), (2.2d), (2.5b), and (2.1r).

The Lagrangean dual problem is:

$$Z_{LD} = \max_{\lambda}(Z_{LD1}(\lambda) + Z_{LD2}(\lambda)) \tag{2.9}$$

Lagrangean decomposition method is based on a special case of Lagrangean relaxation where each set of constraints appears in one of the two subproblems. To improve the speed of convergence, some constraint sets can be added to both subproblems [70, 48]. In our case, we add duplicates of constraints (2.1e) and (2.1f) as constraints (2.8b) and (2.8c) to (SP2).

2.4.2.3 Algorithm for LDH

We develop a heuristic that uses the LD formulation developed in the previous section in order to solve (P). The algorithm begins initializing Lagrangean multipliers λ_{idtk}, the step size σ_{idtk}, ε and the scalar $\nu^n \in (0,2]$. The LB and

UB on the optimal objective value of (P) are also initialized to $-\infty$ and $+\infty$. The maximum number of iterations to be performed is determined by the parameter n^{max}. In step 2, we obtain the lower bound for iteration n, i.e. LB^n, by solving SP1 and SP2 and then calculating the sum of the objective values of these subproblems. In step 3, we find a feasible solution for (P) using the strategy explained in the next section. The total cost of the feasible solution provides an UB on the optimal solution in iteration n i.e. UB^n. Next, we update the LB^{max} if $LB^n > LB^{max}$. We keep track of the incumbent solution by updating the UB^{min} value if $UB^n < UB^{min}$. Here, $LB^{max}(UB^{min})$ is the best lower bound (upper bound) found up to iteration n. The Lagrangean multipliers are updated iteratively via subgradient optimization at the end of each iteration. The scalar ν^n is reduced by a certain amount if the lower bound fails to improve after a fixed number of iterations. The algorithm terminates when one of the following conditions is satisfied: (*i*) the best lower bound is equal to the upper bound found so far as this indicates that an optimal solution is found; (*ii*) the number of iterations reaches the prespecified bound; (*iii*) the scalar ν^n is less than or equal to ε (a predefined number close to zero). The pseudo-code of LDH is presented by Algorithm1.

Algorithm 1 Pseudocode of LDH Algorithm

procedure ▷
Step 1:
Initialize λ_{idtk}, UB^{min}, LB^{max}, σ^{n}_{idtk}, ν^{n}, n^{max}, ε
Step 2:
Solve (SP1) and (SP2)
Calculate $LB^{n} = Z_{LD1}(\lambda^{n}) + Z_{LD2}(\lambda^{n})$
If $LB^{n} \geq LB^{max}$
$LB^{max} = LB^{n}$
End If
Step 3:
Find UB^{n}
If $UB^{n} \leq UB^{min}$
$UB^{min} = UB^{n}$
End If
Step 4:
Update λ_{idtk} using the equations (2.10) and (2.11)
Step 5:
Update ν if required
Step 6:
If $n > n^{max}$ or $\nu^{n} < \varepsilon$ or $UB^{min} - LB^{max} \leq \varepsilon$
Stop
End If
Otherwise **Go To** step 2

The Lagrangian multipliers used in formulation (LD) are updated in Algorithm 1 using a subgradient optimization method. Subgradient optimization is popular among researchers due to its efficiency in solving MIP problems with LD [160, 31]. Readers are referred to [78] for the detailed discussion about computational performance and theoretical convergence properties of the subgradient method. We use this method to update the dual multipliers for iteration $n+1$ if the algorithm does not terminate in iteration n. The Lagrangean multipliers λ^{n+1}_{idtk} for iteration $n+1$ are calculated using the following equation:

$$\lambda_{idtk}^{n+1} = \lambda_{idtk}^{n} + \sigma_{idtk}^{n} * (X_{idtk} - Y_{idtk}) \tag{2.10}$$

where stepsize σ is calculated using

$$\sigma_{idtk}^{n} = \frac{\nu^n * (UB^{min} - LB^{max})}{\sum_{i=1}^{N}\sum_{d=1}^{\delta}\sum_{t=1}^{T}\sum_{k=1}^{K}(X_{idtk} - Y_{idtk})^2}. \tag{2.11}$$

2.4.3 Finding the Upper Bound

In this section, we discuss two methods for generating upper bounds for (P). These methods use solutions to (SP1) to generate feasible solutions for (P). Since formulation (LD) is a relaxation of (P), any feasible solution to (P) is also feasible solution to (LD). Since we have a minimization problem, this solution provides a pessimistic estimate or a valid upper bound for (LD). Proposition 2.4.1 shows that any feasible OR schedule can be used to generate a feasible solution to (P). Our next proposition shows that a solution to (SP1) yields a feasible OR schedule and therefore can be used to generate a feasible solution for (P).

Proposition 2. *If the tuple (X, Z, Q, O, U) is a feasible solution to problem (SP1), then, it is a feasible solution to problem (S) and yields a feasible OR schedule.*

Proof. Model (SP1) contains all the OR scheduling constraints of model (S). Therefore, (SP1) is solved using the same feasible region as (S) with a different objective function. As a result, a solution to an instance of (SP1) is always a feasible solution to (S). □

We use Propositions 2.4.1 and 2 together to generate the upper bounds.

2.4.3.1 Method 1: Simple Upper-Bounding Heuristic (LD-SH)

LD-SH heuristic generates feasible solutions for (P) using feasible solutions to the Lagrangian subproblem (SP1). This heuristic is very similar to heuristic H1 as it is based on fixing all variables of (SP1) in (P). In every iteration of LDH, SP1 generates a feasible OR schedule captured by tuple (X, Z, Q, O, U). We enforce this assignment by fixing the OR schedule as $X = \hat{X}, Q = \hat{Q}, O = \hat{O}, U = \hat{U},$ and $Z = \hat{Z}$ in (P) to initialize (F). Next, we solve (F) to assign instruments to surgeries to minimize the total cost of instruments. LD-SH is used to obtain one feasible solution to (P) in every iteration of LDH. The pseudocode of the heuristic is given below.

Algorithm 2 Pseudocode of Heuristic LD-SH

procedure ▷
  Step 1: Obtain parameters $\hat{X}, \hat{Z}, \hat{Q}, \hat{O}, \hat{U}$ from (SP1) solution
  Step 2: Initialize (F) by enforcing the OR schedule
  Step 3: Solve (F)
  Step 4: Return Z_F to LDH as UB^n

2.4.3.2 Method 2: Bender's Decomposition-Based Upper Bounding Heuristic (LD-BD)

There is a special structure within the MIP formulation of SP1 that allows us to decompose it into smaller and more manageable optimization problems. In particular, we decompose the weekly multi-OR scheduling problem into single day, single OR scheduling problems. Note that the ORs scheduled for a given day do not share any resources with other ORs in SP1. Given a tentative patient to (day, OR) assignment, this renders the scheduling of surgeries for the tuple (day, OR) of one day independent of other (day, OR) tuples of the same day, and a OR scheduling model can be solved for each (day, OR) tuple independently. These independent problems must adhere to the OR scheduling constraints. A weekly schedule is then constructed by combining the schedules of each individual (day, OR) tuple. Once we have a feasible OR schedule, we follow the process outlined in the algorithm 2 to obtain a feasible solution to (P). We apply Bender's decomposition method to generate new surgery to (day, OR) assignments. The procedure is repeated until certain termination criteria are met. We keep track of the best feasible solution, i.e. the incumbent solution for (P). This incumbent solution for the integrated problem is used as UB^{min} in LDH.

As mentioned above, we use Bender's decomposition (BD) to obtain surgery to (day, OR) assignments. In particular, our decomposition method is based on the Logic-Based Bender's Decomposition [79]. We choose this method because surgery assignment to (day, OR) tuples results in MIP subproblems, one for each (day, OR) tuple. In this subsection, we present a BD-based algorithm that decomposes the original MIP model (SP1) into (i) a MIP master problem (MP) that assigns surgeries to (day, OR) tuples to minimize the total cost associated with opening the ORs on each day and (ii) MIP scheduling sub-problems (SP), one for each (day, OR) tuple, that minimizes the cost of overtime, idle-time and the cost of surgery assignments due to λ multipliers. The objective function value of the relaxed problem, which is an approximation of the original problem, is improved at each iteration by adding valid inequalities for each scenario. These are known as optimality cuts. In general, the algorithm uses Bender's feasibility and optimality cuts to communicate the infeasible status or suboptimal solutions of its SPs to the MP. Since it is easier to obtain a feasible assignment of surgeries to a (day, OR) tuple, no feasibility cuts are needed. Thus, all discussions in the sections thereafter will be limited to optimality cuts only. The algorithm is terminated after reaching the iteration limit or when the MP generates an assignment that is optimal to SP1.

Master Problem

The master problem (MP) determines the assignment of surgeries to (day, OR) tuples, and therefore which ORs to open. Thus, we introduce variables $X_{idk} \in \{0,1\}$ that decide the assignment of surgeries to each (day, OR) pair. We define an auxiliary variable θ_{dk} that represents the cost of scheduling decisions made in the subproblems. The (MP) is given below. (MP) uses the same indices and sets as (P) unless stated otherwise.

(MP):

$$Z_{MP} = \min \sum_{d=1}^{\delta} \sum_{k=1}^{K} [c^k Q_{dk} + \theta_{dk}] \tag{2.12a}$$

$$X_{idk} \leq Q_{dk}, \qquad \forall i,d,k, \tag{2.12b}$$

$$Q_{dk} \geq Q_{dk'}, \qquad \forall k,k' \in \mathbb{K}, k' > k, d, \tag{2.12c}$$

$$\sum_{d=1}^{\delta} \sum_{k=1}^{K} X_{idk} = 1 \qquad \forall i, \tag{2.12d}$$

$$\sum_{i=1}^{N} X_{idk} * l_i \leq T, \qquad \forall d \in \mathbb{D}, k \in \mathbb{K}, \tag{2.12e}$$

$$\text{Combinatorial Bender's cuts} \tag{2.12f}$$

$$X_{idk} \in \{0,1\}, \qquad Q_{dk} \in \{0,1\}$$

Constraints (2.12b) ensure that if a surgery is scheduled in an OR, then that OR is opened. Constraints (2.12d) ensure that all surgeries are assigned to only one (day, OR) pair. Similar to (P), symmetry-breaking constraints (2.12c) are added to (MP). Constraints (2.12e) ensure the feasibility of the resulting subproblems. Constraints (2.12f) are the set of optimality cuts added to (MP).

Sub-problem for each (day, OR)

Each iteration of the master problem determines the values of the assignment variables X_{idk}. This results in a known set of allocated patients for each (day, OR). Let the set $\mathbb{N}_{dk}$ represent this allocation and let N_{dk} denote the number of surgeries scheduled in (day d, OR k). Since the assignment of surgeries to a (day, OR) pair is fixed, in the subproblems, the cost of the sequencing decisions is calculated. Model (SP) is used to solve smaller independent subproblems, one for each (day, OR) tuple. Furthermore, the subproblem (SP) for a tuple (d,k) is solved only when the value of $Q_{dk} = 1$ i.e. the OR k is opened on day d.

Model (SP_{dk}) defined for each (d,k) determines the start time for each surgery in the set $\mathbb{N}_{dk}$. Thus, we introduce variables $\bar{X}_{it} \in \{0,1\}$ that denote whether surgery i begins in time slot t or not for all i and t. Variables $\bar{Z}_{it} \in \{0,1\}$

show whether the OR is occupied at time t by surgery i. Variables $\bar{O}$ and $\bar{U}$ determine the overtime and idle-time respectively. Model (SP_{dk}) is given below. This model uses the same indices and sets as (P) unless stated otherwise.

(SP_{dk}):

$$Z_{dk}^{SP} = \min\ c^u\bar{U} + c^o\bar{O} + \sum_{i=1}^{N_{dk}}\sum_{t=1}^{T}\lambda_{it}\bar{X}_{it} \tag{2.13a}$$

$$t * \bar{Z}_{it} \leq s + \bar{O} \qquad \forall t, i \in \mathbb{N}_{dk} \tag{2.13b}$$

$$s - \sum_{i=1}^{N_{dk}}\sum_{t=1}^{T}\bar{Z}_{it} \leq \bar{U} \tag{2.13c}$$

$$\sum_{t'=t}^{t+l_i-1}\bar{Z}_{it'} \geq l_i * \bar{X}_{it} \qquad \forall i \in \mathbb{N}_{dk}, 1 \leq t \leq T - l_i + 1, \tag{2.13d}$$

$$\sum_{t=1}^{T}\bar{Z}_{it} = l_i \qquad \forall i \in \mathbb{N}_{dk}, \tag{2.13e}$$

$$\sum_{i=1}^{N_{dk}}\bar{Z}_{it} \leq 1 \qquad \forall t, \tag{2.13f}$$

$$\sum_{t=1}^{T}\bar{X}_{it} = 1 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2.13g}$$

$$\sum_{t=T-l_i+1}^{T}\bar{X}_{it} = 0 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2.13h}$$

$$\bar{Z}_{it} \in \{0,1\}, \qquad \bar{O} \geq 0, \qquad \bar{U} \geq 0, \qquad \bar{X}_{it} \in \{0,1\} \tag{2.13i}$$

The objective function (2.13a) minimizes the cost of overtime, idle-time, and the cost of surgery assignment. Constraints (2.13b) and (2.13c) calculate the OR overtime and idle-time respectively. Constraints (2.13d) ensure that an OR remains busy during surgery. That is, if surgery i begins in time slot t, then, no other surgery will be scheduled during time slots t to $t + l_i$. Constraints (2.13e) ensure that the number of times lots assigned to a surgery is equal to the duration of the surgery. Constraints (2.13f) ensure that at most one surgery can be scheduled during a particular time slot in a given OR. Constraints (2.13g) and (2.13h) ensure that each surgery is scheduled and completed before the end of the day.

Optimality Cuts

In each iteration of BD, a solution to (MP) is the optimal solution since $Z_{dk}^{SP} = \theta_{dk}\ \forall SP_{dk}$, or provides an upper bound for the model since $Z_{dk}^{SP} >= \theta_{dk}$ for any SP_{dk}. If a solution to (MP) is not optimal, then the following "combinatorial Bender's optimality cut" is a valid inequality:

$$\theta_{dk} \geq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)(\sum_{i \notin \mathbb{N}_{dk}} X_{idk} + \sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk})) \qquad \forall \{(d,k) : \bar{Q}_{dk} = 1\} \tag{2.14a}$$

Where X^*_{idk} represents the assignment of the surgeries in (MP), Z_{dk}^{SP} is the optimal solution to SP_{dk}, and L is a valid LB to θ_{dk}. We refer the reader to [150] for more information on these cuts.

The optimality cut (2.14a) presented above is a valid cut for the feasible region in every iteration. However, the use of these cuts can lead to a very weak formulation, since inequality (2.14a) adds one optimality cut in every iteration for each (day, OR) pair and removes only one solution from future iterations. It is desirable if the cut (2.14a) can be further strengthened. We propose two additional inequalities that can be added to the master problem based on the information obtained from each subproblem.

If adding any surgery to the OR (d,k) increases the current cost, (2.14a) can be further strengthened. In other words, if every superset of the support of X^* yields $\theta_{dk} \geq Z_{dk}^{SP}$ the cut can be strengthened as follows:

$$\theta_{dk} \geq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)(\sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk})). \tag{2.15a}$$

On the other hand, if every subset of the support of X^* yields $\theta_{dk} \geq Z_{dk}^{SP}$ the cut can be strengthened as follows:

$$\theta_{dk} \geq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)(\sum_{i : X^*_{idk} = 0} X_{idk}). \tag{2.16a}$$

The proof of validity of these cuts is provided in the Appendix. To identify if there exists a subset or superset of surgeries for each pair of (d,k) that satisfy the conditions given above, we propose two models (M_1) and (M_2). The details of models M1 and M2 are provided in the appendix. Model (M_1) determines whether there exists a superset of $\mathbb{N}_{dk}$ with a cost less than Z_{dk}^{SP}. For this purpose, (M_1) is solved to schedule all other surgeries in $\mathbb{N}$ in addition to the surgeries in $\mathbb{N}_{dk}$. This is done by replacing the set $\mathbb{N}_{dk}$ with the set $\mathbb{N}$ in the objective function and in constraints. However, to enforce the current assignment, we leave constraints (2.13g) as is. If no surgeries in addition to surgeries from set $\mathbb{N}_{dk}$ are scheduled, it implies that there does not exist a superset of surgeries that reduces the cost of scheduling the surgeries in that (d,k). As a result, the cut (2.15a) is added to (MP). Model (M_1) is given below. This model uses the same notations as (SP).

(M1)

$$Z_{\overline{S}} = \min\ c^u\bar{U} + c^o\bar{O} + \sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{it}\bar{X}_{it} \tag{2.17a}$$

$$\sum_{t=1}^{T}\bar{X}_{it} = 1 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2.17b}$$

$$\bar{Z}_{it} \in \{0,1\}, \qquad \bar{O} \geq 0, \qquad \bar{U} \geq 0, \qquad \bar{X}_{it} \in \{0,1\}$$

Subject to: equations (2.13b) - (2.13f), (2.13h), (2.13i)

The model (M_2) determines whether there exists a subset of $\mathbb{N}_{dk}$ with a cost less than Z_{dk}^{SP}. For this purpose, we replace the constraints (2.13g) from (SP) with (2.18b). This allows (M_2) to schedule a subset of surgeries from the set $\mathbb{N}_{dk}$. If all surgeries from the set $\mathbb{N}_{dk}$ are scheduled, it implies that there does not exist a subset of surgeries that reduces the cost of OR. As a result, the cut (2.16a) is added to (MP). The model (M_2) is given below. This model uses the same notation as (SP).

(M2)

$$Z_{\underline{S}} = \min\ c^u\bar{U} + c^o\bar{O} + \sum_{i=1}^{N_{dk}}\sum_{t=1}^{T}\lambda_{it}\bar{X}_{it} \tag{2.18a}$$

$$\sum_{t=1}^{T}\bar{X}_{it} \leq 1 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2.18b}$$

Subject to: equations (2.13b) - (2.13f), (2.13h), (2.13i)

Proposition 3. *Each iteration of Bender's Decomposition generates a feasible solution to (SP1).*

Proof. The model (MP) determines the surgeries scheduled in each (day, OR) tuple. These schedules are feasible for each (day, OR) tuple. Model (SP) uses these assignments to generate a feasible schedule. Therefore, a feasible solution for problem (SP1) can be constructed by combining the schedules of all sub-problems. □

LD-BD heuristic is very similar to the heuristic H1, as it is based on fixing all variables of (SP1) in (P). Each iteration of BD generates a feasible schedule captured by the tuple (X, Z, Q, O,U). We enforce this assignment by fixing the OR schedule as $X = \hat{X}, Q = \hat{Q}, O = \hat{O}, U = \hat{U}$, and $Z = \hat{Z}$ in (P) to initialize (F). Next, we solve (F) to generate an instrument assignment that minimizes the total instrument usage and instrument rental costs. LD-BD obtains one feasible solution to (P) in every iteration of BD. The best feasible solution obtained is used as UB^{min} in LDH. The advantage of using LD-BD is that it can generate several feasible solutions within one iteration of LDH. This method uses a partially guided search by using the λ multiplier values obtained in each iteration of LDH. However, it solves several optimization problems in order to generate the necessary cuts.

Algorithm 3 Pseudocode of Bender's decomposition-based upper bounding heuristic (LD-BD)

procedure ▷
  Step 1: Initialize parameters n, n^{max},ε, LB, UB
  Step 2: Initialize (MP)
  Step 3:
  while $n < n^{max}$ **do**
    Solve (MP) to obtain X^*
    for $d \in \delta$, $k \in K$ **do**
      Initialize (SP)
      Solve (SP) to compute Z_{dk}^{SP}
    end for
    Update LB = Z_{MP}
    if $Z^{MP} - \sum_{dk} \theta_{dk} + \sum_{dk} Z_{dk}^{SP} \geq UB$
      Update UB
    Solve (F) to compute Z_F
    if $(UB - LB)/UB < \varepsilon$
      return UB, X
    if $(Z_F - LB^{max})/Z_F < \varepsilon$
      return Z_F, X
    elif $\theta_{dk} < Z_{dk}^{SP} - \varepsilon$
      develop and add necessary Bender's cuts
    end if
  end while

The pseudocode of the LD-BD algorithm is presented by Algorithm 3. The algorithm begins by initializing the parameters such as iteration counter n to zero, the maximum number of iterations to n^{max}, LB and UB on the optimal solution of the SP1 to $-\infty$ and $+\infty$. In step 2, we initialize MP. In step 3, MP is solved to optimality to obtain surgery assignment X^* and objective function value Z_{MP}. Z_{MP} provides a LB for (SP1). Next, we fix the surgery assignments X^* and solve SP_{dk}. The solution of each SP_{dk} is used to construct a feasible solution for SP1. This solution is used to update UB if $Z^{MP} - \sum_{dk} \theta_{dk} + \sum_{dk} Z_{dk}^{SP} \geq UB$. For each schedule obtained, we solve (F) to obtain Z_F. We terminate the algorithm LD-BD and LDH if $Z_F = LB^{max}$, where LB^{max} is the best lower bound found for the LDH so far, since $Z_F = LB^{max}$ indicates that an optimal solution to (P) was found. Depending on the feedback, combinatorial bender's cuts are added to the (MP). Step 3 is repeated until the algorithm is terminated. The algorithm terminates when the error gap is less than ε as this indicates that an optimal solution for SP1 is found or when the maximum number of iterations, n^{max},is reached.

2.5 Computational Results

In this section, we first discuss the data we use to generate realistic problem instances for our computational experiments. Next, we compare the performance of the proposed algorithms. Finally, we estimate the value of integrated OR scheduling and assess the impact of critical factors via sensitivity analysis to address the research questions outlined in Section 2.1.

2.5.1 Parameter Estimation

The research presented in this chapter was conducted in collaboration with GMH, one of the seven campuses of Prisma Health in South Carolina, USA. GMH provides general inpatient services and specialized treatments for heart disease and cancer. The hospital also houses the Family Birthplace, the Children's Hospital, and the Children's Emergency Center. The research team collaborated with the Perioperative Services Department (PSD), which oversees the OR scheduling, sterilization, and inventory management processes. The PSD is also responsible for transporting materials to and from the ORs. For this case study, we assume that the material handling delay is constant and included in the data. We obtained a data set from GMH that includes information on the type of surgery, the scheduled time, the scheduled duration, and the actual duration of a surgery. GMH offers 46 different types of surgical services. Our experimental analysis focuses on 3 types of surgeries; ENT, Orthopedic, and Neurology. We focus on these surgical services because the data set shows that they are scheduled multiple times per day. Furthermore, Orthopedic and Neurology surgeries generally require more expensive instruments, and currently use rental instruments to meet demand. Therefore, there is an opportunity to reduce the cost of instruments through reuse. The surgical cases to be scheduled in the ORs were randomly sampled from the data. The number of scheduling slots required for each surgery in our computational experiments was also calculated based on the actual duration of the surgeries obtained from the data. To convert the duration of surgeries to time slots l, we round the actual duration to the nearest half-hour. We solve the problem to generate a weekly OR schedule, and hence the number of days, δ is chosen to be 5 days. Our data analysis indicates that different services do not share more than 3 ORs. Therefore, the number of ORs, K, is set to 3 ORs. The number of surgeries to be scheduled is set to $N = 30$. This number is also inspired by the data that indicate that up to 2-3 surgeries are scheduled in an OR in a typical day. We consider instruments of six different types, i.e. $R = 6$, and the inventory level, I_r, is set to 2 instruments per type. The requirement matrix for surgical instruments was generated randomly for each type of surgery. We assume that surgeries may not require more than 1 instrument of each type. This is a reasonable assumption, since we focus primarily on surgical instruments that are very expensive to purchase. From the data collected from GMH, we observed that instances of multiple expensive instruments of the same type being used for one surgery are rare. One can consider expensive instrument sets instead of individual instruments, as it is also common practice to rent an entire set for a surgery.

Based on an assessment of various cost criteria, we set the following parameters: fixed cost of opening the OR, c^k = \$360, session length $s = 16$, total number of time slots $T = 20$, overtime cost, c^o =\$100, and the idle time cost, c^u =

\$22.5. We assume that using 2 hours of overtime is equivalent to opening a new OR. We set the usage cost as $c^F = \$18$ and the instrument rental cost as $c^R = \$36$. The OR opening cost, overtime cost, and idle time costs are kept the same for all experiments. The sterilization time, γ, is set to 6 time intervals, that is, 3 hours, based on an expert opinion in GMH. We assume that the instruments are sterilized immediately after surgery. As the sterilization department works for 24 hours in shifts, surgical instruments can be sterilized overnight and it is assumed that they are available to use at the start of the next day.

2.5.2 Computational Performance of the Proposed Methods

We analyze the performance of the proposed algorithms. The proposed algorithms are coded using the Python programming language and are executed on a high-performance HP computer with an Intel Xeon processor with 16 cores and 62GB of memory. We distinguish between small, $N = 10$, medium, $N = 20$, and large, $N = 30$ problem instances. Each set of problems consists of 10 different instances and is solved via the Gurobi optimization solver using model (P). We solve these problem sets again by using the proposed approaches, i.e. LD-SH, and LD-BD algorithms. In our implementation of the LDH, we obtained ideal parameter values via extensive experiments. These values are then used for all the problems in this section. We begin with $\lambda = 1$ and $\nu = 1.8$. The value of ν is reduced by 20% if there was no improvement in the last 5 iterations of LDH. LDH is terminated after reaching $n^{max} = 50$ iterations or if the time limit of 5 hours is reached. The algorithm is also terminated if the error gap is less than 1%. We use the same parameters for both LDH-based algorithms. All three algorithms were stopped after 5 hours. When problem (P) is solved using Gurobi, both upper and lower bounds are provided by the Gurobi solver. When problems are solved by our proposed solution approach, the lower bound is obtained from the LD. The error gap for these experiments is calculated as follows.

$$\text{Error Gap}(\%) = \frac{UB - LB}{UB} * 100 \tag{2.19}$$

The results for these experiments are summarized in Tables 2.2, 2.3, and 2.4 for small, medium, and large problem instances, respectively. We make the following observations.

Observation 1: For small instances, the bounds provided by LD-SH and LD-BD are identical and, in some cases, better than those provided by Gurobi. It can be observed that the LD-SH and LD-BD methods find solutions within a 1% error gap very quickly. This is because the number of variables and constraints in our problem depends on the number of surgeries. See Table 2.2.

Observation 2: For medium size problems, bounds provided by LD-SH and LD-BD are identical and significantly better than Gurobi. Both LDH-based methods obtain solutions within a 2% gap on average. See Table 2.3.

Observation 3: For large-size problems, bounds provided by LD-SH and LD-BD are close and significantly better than Gurobi. Both LDH-based methods obtain solutions within a 7% gap on average. Gurobi does not find solutions within a 1% optimality gap in 5 hours for any of the large or medium problems. See Table 2.4.

	Problem (P) Gurobi		LD-SH		LD-BD	
	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)
Avg	0.39%	1.76	0.00%	0.01	0.08%	0.08
Min	0.00%	0.30	0.00%	0.01	0.00%	0.01
Max	3.91%	0.00	0.08%	0.02	0.81%	0.30

Table 2.2: Experiments for Computational Performance (Small)

	Problem (P) Gurobi		LD-SH		LD-BD	
	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)
Avg	7.31%	5.00	1.60%	1.10	1.80%	2.20
Min	1.59%	5.00	0.00%	0.10	0.00%	0.10
Max	14.30%	5.00	8.07%	5.00	9.39%	5.00

Table 2.3: Experiments for Computational Performance (Medium)

	Problem (P) Gurobi		LD-SH		LD-BD	
	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)	Error Gap	Run Time (Hours)
Avg	11.88%	5.00	5.61%	3.90	6.21%	4.22
Min	6.89%	5.00	0.34%	0.37	0.41%	0.66
Max	15.40%	5.00	10.22%	5.00	10.22%	5.00

Table 2.4: Experiments for Computational Performance (Large)

These results show that our decomposition approach is effective, since the resulting algorithms provide better bounds in reasonable time in most cases when compared with solving (P) directly with Gurobi. The performance of LD-BD can be improved by developing stronger optimality cuts or by running the sub-problems of BD in parallel. The advantage of LD-BD is that it generates several feasible OR schedules, one in each iteration.

Several researchers have highlighted the effect of the ratios of fixed cost f to variable cost v on the difficulty of solving the problem and therefore the quality of the solution for fixed-charge cost functions [118, 87]. To further analyze the performance of the proposed algorithms, we conduct sensitivity analyses with respect to f/v ratios. For these experiments, we vary the rental instrument cost c^R from 18, 36, 90, 180, 360, 450 to generate ratio values (r_1 - r_6) as 0.05, 0.1, 0.25, 0.5, 1, 1.25 respectively. Figure 2.4 summarizes the results of these experiments.

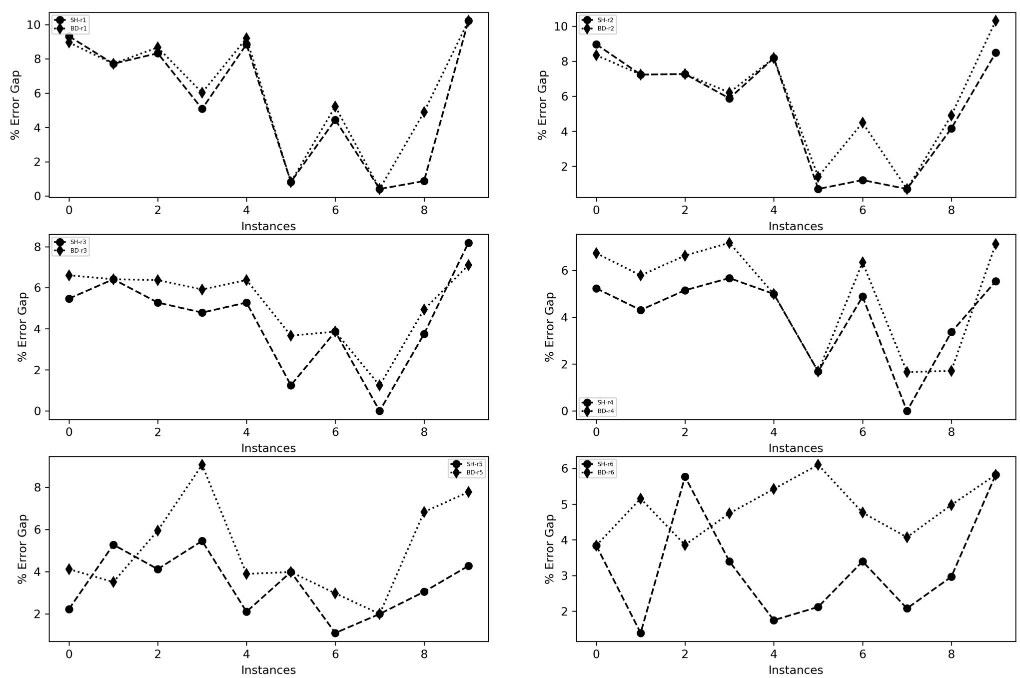


Figure 2.4: Sensitivity Analysis Day vs OR Instrument costs

The results indicate that for small values of cost ratios,i.e., $r_1 - r_4$ the performance of both algorithms is comparable. For higher values of cost ratios,i.e., r_5 and r_6 the performance of LD-BD is inferior to LD-SH. For higher values of cost ratio values, the problem is easier to solve with LD-SH. As LD-BD algorithm spends more time solving a series of MIP problems as part of the BD approach, fewer iterations of the Langrangian decomposition are performed in the same time. Therefore, the results of LD-BD algorithm are inferior for higher cost ratio values.

2.5.3 Discussion of Results

Research question 1: ***How do integrating decisions about OR scheduling and inventory management impact the OR costs?***

To address this research question, we implement the construction heuristic H1, which initially solves model (S) and then uses the corresponding OR schedule to assign instruments to surgeries, that is, model (F). Note that this follows the current practice at GMH, where the OR schedule is generated first, and then instruments are assigned to ORs based on the schedule. Next, we solve the integrated OR scheduling problem (P). The objective function value of H1, Z_{H1}, is compared with the objective function value of problem (P), Z_P. We define *improvement potential* as the percentage improvement in total costs if a hospital uses integrated OR scheduling instead of the current practice of OR scheduling and inventory assignment. The improvement potential for these experiments is calculated as follows.

$$\text{Improvement Potential}(\%) = \frac{Z_P - Z_{H1}}{Z_P} * 100 \tag{2.20}$$

We conducted additional experiments to address research question 1. We distinguish between short, $l < 4$(hours), and long $l > 4$(hours) surgeries. These experiments allow us to examine under what conditions the integrated OR schedule performs well when compared to the current schedule. The attributes of problem sets are summarized in Table 2.5. Each problem set consists of 10 different problem instances. Tables 2.6, 2.7, 2.8 summarize the results of these experiments.

Characteristics of Problems			
Surgery Duration	Size		
	Small	Medium	Large
Short	Problem 1	Problem 2	Problem 3
Mix of short and long	Problem 4	Problem 5	Problem 6
Long	Problem 7	Problem 8	Problem 9

Table 2.5: Problem Setup

We present the minimum, maximum, and average improvement potentials for each problem. We also calculate the difference between the number of instruments rented in H1 compared to problem (P). We make the following observations.

	Improvement Potential			Rental Instrument Reduction		
Problems	Minimum	Average	Maximum	Minimum	Average	Maximum
Problem 1	0%	3%	5%	0.0	2.0	4.0
Problem 2	0%	3%	7%	0.0	4.3	9.0
Problem 3	1%	2%	7%	1.0	4.0	13.0

Table 2.6: Result of Model Validation for Short Surgeries

	Improvement Potential			Rental Instrument Reduction		
Problems	Minimum	Average	Maximum	Minimum	Average	Maximum
Problem 7	0%	1%	2%	0.0	0.6	3.0
Problem 8	0%	2%	8%	0.0	1.4	4.0
Problem 9	0%	3%	6%	1.0	2.4	5.0

Table 2.7: Result of Model Validation for Long Surgeries

	Improvement Potential			Rental Instrument Reduction		
Problems	Minimum	Average	Maximum	Minimum	Average	Maximum
Problem 4	0%	1%	5%	0.0	0.8	4.0
Problem 5	0%	1%	2%	0.0	1.9	3.0
Problem 6	0%	1%	2%	0.0	2.5	5.0

Table 2.8: Result of Model Validation for Mixed Duration Surgeries

Observation: For short-duration surgeries, we observe that fewer instruments are rented. In short-duration surgeries, 2 to 4.3 fewer instruments were rented on the average, as compared to current practice. This is because many short-duration surgeries are scheduled in a day. Thus, instruments are reused frequently. See Table 2.6. For long-duration surgeries, the number of rented instruments reduced slightly. We observe that the average number of rented instruments is 0.6 to 2.4 units lower than the current practice. This is mainly because only a few long-duration surgeries are scheduled in a day, thus, instruments are reused sporadically. Therefore, integrated OR scheduling demonstrates lower improvement potential when scheduling long-duration surgeries. See Table 2.7. The observed reduction in the number of rented instruments, when scheduling a mix of surgeries with short and long duration, is consistent with previous observations. We observe that the average number of rented instruments 0.8 to 2.5 units lower than the current practice. The daily requirements for rental instruments are lower than those of short-duration surgeries and more than those of long-duration surgeries. See Table 2.8.

The result indicates that there is considerable value in coordinating instrument inventory decisions with OR schedules. This is true, especially for hospitals that reserve ORs to schedule shorter surgeries such as dental or ENT surgeries. These services schedule more surgeries and have higher instrument requirement per day. Therefore more instruments can be reused in a day. Therefore, the highest improvement can be achieved when surgeries have short duration ($l <$ 4hours). The benefits of using the proposed integrated OR scheduling and inventory model increase with the number of surgeries scheduled, which could be due to integrating the scheduling of many ORs, or due to increasing the planning period. Note that, Heuristic H1 is designed to minimize the cost of OR using optimization solvers. In addition, for any given schedule, Heuristic H1 also assigns the surgical instruments optimally. At GMH and other hospitals, this is usually not the case. Therefore, the value of integrated OR scheduling could be higher than that observed in our experiments. Moreover, this integration can lead to reduction in instrument-related costs and improved utilization of surgical instrument inventory. It also improves healthcare access for patients because hospitals can schedule more add-on surgeries using the same number of rental instruments/ inventoried instruments.

Research question 2: ***How does the inventory level of surgical instruments, the length of planning horizon, and number of ORs scheduled in parallel impact the system utilization and the cost of an integrated OR schedule?***

To address this research question, we analyze how sensitive the integrated scheduling problem is to changes in problem parameters by changing one parameter at a time. In particular, we vary the length of the planning horizon (in days), number of ORs working in parallel, and the inventory level of surgical instruments to investigate how the availability of these resources impacts the cost of ORs and the schedule itself. We consider problem instances with a large size, i.e. $N = 30$, and surgeries with mixed-duration. In our experimental design, we first change the level of inventory from 1 instrument to 3 instruments by increasing it by 1 unit each time. For each inventory level, we choose the unique combination of number of days from $\delta = \{3,5\}$ and number of ORs from $K = \{3,5\}$. The experiment setup is outlined in Table 2.9. Each problem consists of 10 instances. The instances are solved using the LD-SH algorithm described in section 2.4.

Characteristics of Problems		
Number of Days	Number of ORs	
	3	5
3	Problem 1	Problem 2
5	Problem 3	Problem 4

Table 2.9

Figures 2.5 and 2.6 summarize the results of our experiments. The boxplots in Figure 2.5, depict the number of overtime slots used and the number of idle-time slots in each configuration. The boxplots in Figure 2.6, depict the number of instruments used, rented, and reused in each configuration. The first 6 box-plots denote the combinations with 5 ORs, while the last 6 box-plots denote the combinations with 3 ORs. We make the following observations.

Observation 1: The OR-related costs depend on the total available OR time, i.e. the number of days, δ and the number of ORs, K. Increasing OR time provides more flexibility in scheduling surgeries. Therefore, the scheduler can leverage the trade-off between using idletime, overtime or using rental instruments. The results indicate that the OR costs depend only on the total available OR time. When δ and K are set to the lowest value, on average 13 units of overtime were used. Since the available OR time is sufficient for all other combinations, there is little to no overtime used. In those cases, a higher idle-time cost was observed.

Observation 2: The number of instruments used, reused and rented depends on the total available OR time. As the number of days are increased while keeping the same number of ORs, number of instruments used from inventory increases. This increase is more when compared to increasing the number of ORs while keeping the same planning horizon. This is because when planning horizon is longer, the scheduler has more flexibility to utilize more instruments by scheduling surgeries over multiple days. The rate of increase in instrument usage reduces as the level of inventory is increased. In smaller planning horizon, the number of instruments reused is higher when compared to larger planning horizon. This is because all the surgeries must be scheduled by opening more ORs. Therefore, the model schedules the surgeries to minimize the use of rental instruments. Finally, as the number of days increased, fewer rental instruments are used.

Observation 3: As the level of inventory increases, it is not necessary to reuse the instruments. Therefore, with increasing inventory levels, a reduction in reused instruments was observed. Similarly, increasing the level of inventory increases the number of instruments used from the inventory. This increase in the used instrument diminishes more when the level of inventory is increased from 1 unit to 2 units. As a consequence of the increased utilization of inventoried instruments, fewer instruments are rented.

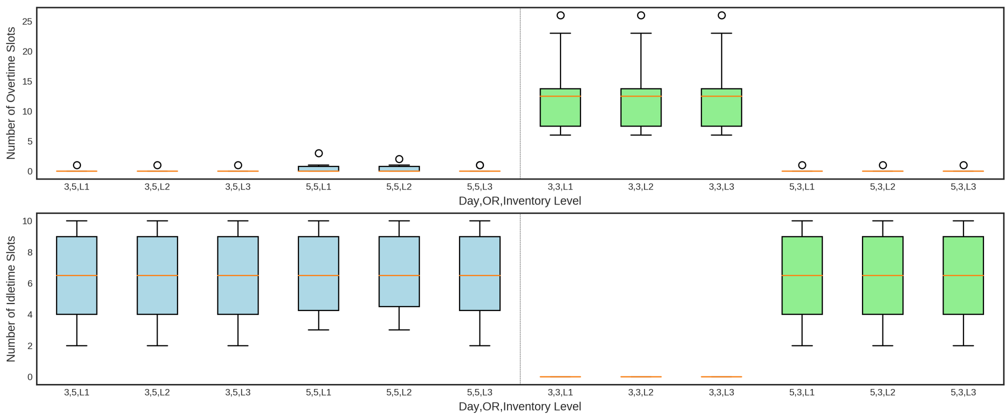


Figure 2.5: OR costs

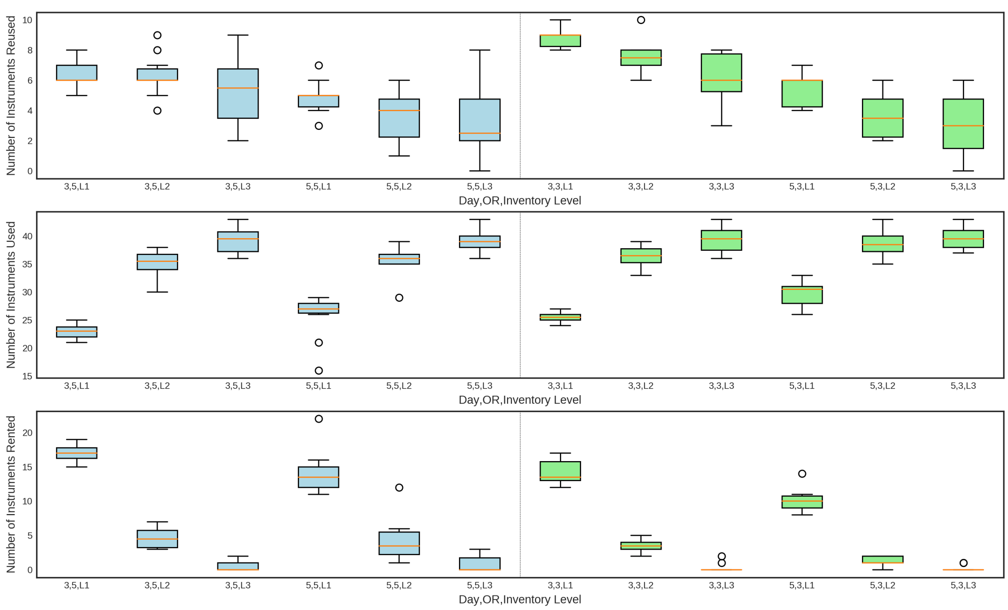


Figure 2.6: Sensitivity Analysis Day vs OR Instrument costs

2.6 Conclusion

The management of rented surgical instruments has recently garnered the attention of researchers. Hospitals avoid purchasing a large number of surgical instruments due to their high costs. In addition, surgeons may request newer and technologically more advanced instruments, making on hand inventory quickly obsolete. Hospitals frequently supplement their inventory via rental instruments. However, these decisions are often not re-evaluated. As a result, hospitals end up with a large number of rental instruments on their shelves. Renting an instrument requires (*i*) processing of related paperwork, (*ii*) following sterilization procedures for rental, which are often different that owned instruments, and (*iii*) paying vendors upon use of the rental instruments that are stored in hospitals' own storage space. The data obtained from GMH indicates that GMH uses rental instruments frequently. The models presented in this chapter are motivated by the opportunities for improvement observed in GMHs inventory management and OR scheduling, which led to the following research questions: *How do integrating decisions about OR scheduling and inventory management impact the OR costs?* and *How does the inventory level of surgical instruments, the length of planning horizon, and number of ORs scheduled in parallel impact the system utilization and the cost of an integrated OR schedule?*

In order to address these research questions, we propose a new approach to OR scheduling that coordinates scheduling decisions with decisions about the management of surgical instrument inventory. We propose an integrated OR scheduling model that jointly determines OR schedules and the assignment of surgical instruments to scheduled surgeries. This model identifies an OR schedule with the objective of minimizing the cost of opening the ORs, overtime, and idle-time along with the costs of using and renting instruments. The results of our model were compared with the approach currently used in OR scheduling where the instrument assignments are determined after the OR schedule is set. Our computational experiments indicate that significant cost improvement and reduction in the number of rental instruments required can be achieved with coordinated decision-making. This is especially true for scheduling of shorter surgeries. When surgery duration is short, our experiments showed a high improvement potential as well as a higher reduction in the number of instruments rented. To solve the integrated OR scheduling model efficiently, we propose an easy-to-implement construction heuristic and a Lagrangean decomposition-based heuristic. We propose two approaches to generate valid upper bound for the optimal solution; a simple heuristic and a Benders decomposition-based heuristic. Both approaches do determine the OR schedule before making the instrument assignment decisions. A subgradient method is then used to iteratively find better upper and lower bounds on the optimal solution. The Lagrangean decomposition-based heuristic is terminated when optimal solutions are found or pre-specified time limit has reached. We compare all solution methods in terms of objective function value and running time. In addition, we also conduct a sensitivity analysis based on the size of the problem, the duration of surgeries, the level of inventory, the number of ORs, the length of the planning horizon, and the cost ratios. Our results indicate that the proposed solution approaches outperform Gurobi, a commercial solver, in terms of running time and solution quality. The results also indicate that both decomposition-based approaches provide high quality bounds.

Based on our experience, hospitals pay more attention to OR scheduling than instrument assignment decisions. This is because major portion of hospitals income is driven by surgical procedures. As a result, inventory management decisions follow the advance OR schedules in conventional approaches. However, our experiments show the cost of ORs can be greatly reduced by integrating these decisions. Hospitals should identify opportunities to coordinate OR scheduling and instrument assignment decisions.

Chapter 3

Metaheuristics

3.1 Introduction

During the past few decades, advances in computer science, artificial intelligence, and operations research have resulted in exact optimization and heuristic methods that have applications for complex decision-making problems. One such method, Tabu Search (TS), forms the basis of this chapter. The Tabu search is a versatile heuristic that is used to solve several combinatorial optimization problems. The objective of Chapter 3 is to develop quick and efficient algorithms to solve the integrated OR scheduling and inventory management problem, and to generate optimal/near-optimal solutions that increase the efficiency of GMH operations. In Chapter 2, we introduced the integrated OR scheduling problem (P), which is a combinatorial optimization problem. As such, the problem is challenging to solve using commercial solvers. To solve this problem efficiently, we proposed a construction heuristic (H1) and a Lagrangean decomposition-based heuristic (LDH) that determine lower bounds (LB) for the optimal objective function value. In addition, we propose two heuristics to determine upper bounds (UB). These methods generate feasible solutions that show promising results for smaller problems. However, they do not work well to solve larger problems because they are iterative procedures and computationally expensive. These challenges have motivated the development of metaheuristics to solve OR scheduling problems, which have been shown to be very effective in solving other combinatorial problems in general [63], and scheduling problems in particular [65, 164].

An interesting characteristic of combinatorial problems is the presence of locally and globally optimal solutions. There can be many solutions that can be optimal locally. Searching for improving solutions in a *neighborhood* can be very useful for combinatorial optimization problems. A neighboring solution of a solution X, is composed of solutions that can be reached from the current solution X with a certain perturbation. For example, if a solution for our problem is a sequence of surgeries $X = \{1,2,3,4\}$, then a solution $\bar{X} = \{2,1,3,4\}$ can be reached by swapping 2 surgeries 1 and 2 in the sequence. Therefore, the solution $\bar{X}$ is in the neighborhood of the solution X. Metaheuristics use a local search approach to intensify the search procedure. Local search heuristics begin with an initial solution and then proceed from the local optimum within one neighborhood to the next neighborhood. At the end of the algorithm, the best local optimum solution is considered the final solution. However, a major disadvantage of a local search procedure is that they are prone to get stuck in a local optimum. If no improving solutions are found in the next neighborhood, the local search heuristic reverts back to the last local optimum found and searches the same neighborhood again and again. The algorithm terminates with a local optimum [116]. These locally optimal solutions can be very far from the globally optimal solutions. The efficiency of the local search is affected by the size of the neighborhood and the mechanism used to determine the neighboring solution to move to. Therefore, techniques to overcome local optimality have been investigated by several authors [157]. Local optima can be escaped by including certain rules in the search procedure to temporarily select a solution even if it is a non-improving one, to have the ability to find improving solutions in the long-run.

TS is an intensive local search method that was first introduced by [63] in 1986. TS uses memory and search history to keep track of recently visited solutions and avoids revisiting the same local solutions repeatedly [164]. The algorithm begins with an initial feasible solution that is iteratively improved. At every iteration, the TS moves to another solution in a defined neighborhood of the current solution. Generally, the best solution in the neighborhood is selected for the move; however, this solution does not have to be an improving one. To avoid being trapped in a local optimum, a Tabu list is used. This list uses short-term memory to store recent moves in the search process. This information is used to guide the next phase of the search process. For example, it is forbidden to repeat a move for a specified number of iterations of the search process. This number is known as the Tabu tenure, and forbidden moves are called Tabu. When the tenure of each Tabu move is over, the move is removed from the Tabu list. The aim is to prevent repeated visits to the same solution, which would limit the search to a certain area of the solution space. A Tabu move can only be made when it is an improvement on the best solution found so far; this condition is called the aspiration criterion [117].

We have two main motivations for developing new algorithms. First, we have developed the H1 construction heuristic and the LD-SH heuristic to solve the integrated OR scheduling problem, problem (P). For large-sized problems, the solutions obtained by LD-SH indicate an error gap of about 6% on average. The solutions obtained by the H1 heuristic are within $< 10\%$ of those of Lagrangian UBs as shown in Tables 2.6, 2.8, and 2.7. This indicates that there is an opportunity to improve the H1 heuristic and the LD-SH in terms of solution quality. Furthermore, LD-SH takes several hours on average to find high-quality solutions. Each iteration of the LD-SH sometimes takes several minutes.

Furthermore, the heuristic H1 requires one to solve two difficult mixed-integer programming models (Models S and F) one after the other. As a result, there is also an opportunity to improve both the H1 heuristic and LD-SH in terms of running time. In light of these observations, we develop a new construction heuristic H2 to find an initial solution. The H2 heuristic obtains OR schedule and instrument assignments in less than 1 second on average for large-sized problems. Second, we expect that using a local search will help us improve this solution quickly; therefore, we develop a TS algorithm that includes a local search. Although TS does not guarantee optimality, they have proven to be fast and efficient in generating near-optimal solutions for difficult scheduling problems [65] and, more recently, this method has been used several times successfully to solve OR scheduling problems [110, 7, 131, 88].

The rest of this chapter is organized as follows: Section 3.2 reviews the literature that is relevant to this work. Section 3.3 provides a detailed description of the algorithm. Section 3.4 introduces a case study and discusses the results of computational experiments. Finally, Section 3.5 summarizes the key takeaways and presents concluding remarks.

3.2 Literature Review

The literature classifies metaheuristics into two categories, (i) single-solution metaheuristics, where a single solution (and search trajectory) is considered at a time, and (ii) population metaheuristics, where multiple solutions can be developed concurrently [61]. Commonly used single-solution metaheuristic methods are Greedy Randomized Adaptive Search Procedure (GRASP), Simulated Annealing (SA), Variable Neighborhood Search (VNS), and TS. Commonly used population metaheuristics are Ant Colony Optimization (ACO), Evolutionary Algorithms (EAs), and Scatter Search (SS). These methods can be distinguished on the basis of how the solution is constructed, i.e. from scratch or via improving the current solution. In the review by [157], metaheuristic methods such as TS and GA are discussed to solve optimization problems. Their discussion highlights the importance of a greedy initial solution in the development of TS.

TS has been shown to produce optimal and near-optimal solutions to a wide variety of scheduling problems, such as personnel scheduling, production scheduling, and scheduling in healthcare while efficiently utilizing computational resources and time.. For example, the work by [65] investigates the employee scheduling problem with the TS algorithm. The integer programming formulation of this problem had one to four million variables. TS algorithm obtained solutions with less than 2% optimality gaps in 20 minutes. TS has numerous applications in the manufacturing setting, for example, in job-shop scheduling [36, 141]. For example, [36] consider a job-shop scheduling problem with two neighborhood structures that swap jobs on critical paths. They find optimal solutions for many instances of the benchmark problem they consider. [47] also used a TS algorithm to solve a flow-shop scheduling problem. Their implementation of TS used three different neighborhood structures (2-swap, 3-swap, and insertion) to diversify the search process. They observed that this TS algorithm obtained solutions as good as the neuro-TS and ACO algorithms when compared. The running time of their TS and ACO algorithms was identical, but TS provided solutions closer to optimal solutions. TS has several applications in parallel machine scheduling problems [24, 96, 77, 117]. For example, [77] solves the problem of unrelated parallel machine scheduling with sequence and machine-dependent setup times to minimize the makepan. Their proposed Tabu Search algorithm uses two perturbation schemes: intra-machine perturbation, which optimizes the sequence of jobs on the machines, and inter-machine perturbation, which balances the assignment of the jobs to the machines. [96] solve similar problem with the insert move in addition to the ones used by [77]. An insert move is a move that inserts a job into any machine. A variant close to our problem is studied by [117]. This problem is a parallel machine scheduling problem with instrument availability constraints. Their implementation of TS shows that the algorithm obtains near-optimal results. They found that the quality of the solution was not affected due to the problem size; however, the large instances of problems required a longer time to find good solutions. The largest problem they solve has 15 jobs, 2 machines, and 8 different types of instrument. The objective of their work is to minimize the makespan. Even though our problem has somewhat similar structures, our work differs because we include several additional decisions to minimize instrument-related costs and OR-related costs. In addition, we also solve a larger-sized problem. Similarly to their implementation, to diversify search process, we restart with a new initial solution in the diversification phase.

Work by [110] reviews the literature related to the metaheuristic for the OR scheduling problem. Similarly to our research, most articles reviewed by [110] develop metaheuristics to solve the *open OR scheduling* problem. They find that TS, GA and ACO are the most commonly used metaheuristics for these problems. TS is the most commonly used single-solution metaheuristic to solve OR scheduling problems. More recently, TS algorithm has been used more successfully to solve OR scheduling problems [11, 95, 7, 151, 88]. For example, [50] propose a TS algorithm and fastest ascent local search method (FALS) for the OR scheduling problem. They use a pairwise exchange mechanism to generate neighborhood solutions. They found that TS algorithm provides high-quality solutions in reasonably short computation times compared to the FALS method. [151] presents TS based metaheuristic to solve OR scheduling problem with the objective of minimizing the make-span. They compare this method with SA and find that TS-based methods provide better quality solutions faster. [131] propose a TS with a discrete event simulation model to minimize patient waiting time, completion time, and number of cancellations. They also propose a Tabu search algorithm that uses integer and binary linear programming to generate initial solutions. To efficiently search the solution space, they use a concept called *candidate list* that evaluates promising solutions first before evaluating other solutions. They find that mathematical programming-enhanced TS algorithm show better results compared to the traditional TS algorithm.

The main difference between our TS algorithm and other tabu search algorithms developed for the OR scheduling problem is that other algorithms use a single neighborhood structure. Our TS algorithm diversifies the search for a better solution using three different neighborhood structures. In addition, for each neighborhood, we build a candidate list separately that consists of promising neighbors so that the neighborhoods are searched efficiently.

There are numerous applications of TS used to make OR scheduling decisions coupled with constraints on other factors such as surgeon availability, patient preferences, limited surgical and recovery resources, etc. For example, [35] solve the OR scheduling problem to minimize the makespan. They also take into account patient sequence-related constraints and other resource-related constraints. [7] propose a Tabu search heuristic to maximize OR utilization and minimize idle time in the neurosurgery department. They solve a problem as large as 5 days, 3 ORs, 10 surgeons, and around 130 cases per month. The solutions obtained using the TS algorithm were within 8% of the solutions obtained by MIP formulation. Work by [95] to solve the OR scheduling problem with the objective of balancing the occupancy in the recovery room after surgery. Their implementation of the TS algorithm provides better results than other local search algorithms. [88] integrate OR scheduling with surgeon schedules. The objective of their work is to minimize overtime costs in hospitals by developing realistic OR schedules. The features of their problems include a significantly large number of surgeries, several types of services that must be scheduled in the ORs, and limited staffing resources. We note that our work extends the applications of TS for OR scheduling problems by that coordinates instrument assignment decisions with OR scheduling.

In the next section, we describe the details of the implementation of our TS algorithm to solve the integrated OR scheduling problem.

3.3 Elements of the Tabu Search approach

3.3.1 Initial Schedule

As most metaheuristic methods begin with an initial solution, we develop a fast and efficient constructive approach to yield an initial solution. Our construction heuristic H2 works in two phases. In phase 1, we generate surgeries for (Day, OR) assignments following a rule of thumb based on the results of the experiments performed in Chapter 2. First, we fix the number of days and the number of ORs to use. Because instruments can be sterilized overnight, an intuitive idea is to schedule surgeries on separate days. Therefore, a list of (Day,OR) tuples is created. The (Day,OR) tuples are used in the day-first order. For example, (Day 1, OR 1) will be used first. After that, (Day 2, OR 1) will be used instead of (Day 1, OR 2). A dictionary is maintained that keeps track of the total duration of surgeries assigned to each (Day,OR). Next, we select the first surgery from the randomly ordered list of surgeries. The surgery is assigned to the first (Day,OR) that is *eligible* for assignment. The surgeries assigned to each (Day,OR) are sequenced in the order they were assigned. This procedure determines the start and end time of each surgery. Once a sequence has been generated, OR related costs such as overtime cost, idle-time cost, and the cost of opening the ORs are computed. The procedure is repeated to identify the number of (Day,OR) to use. The combination with the lowest OR-related cost is chosen as the starting solution of phase 1.

Definition 1. *Let E_{dk} be the end time of surgeries on (Day d, OR k). Let l_i be the duration of the surgery i. Let SL be the planned session length and T be the total number time slots. Then, a (Day, OR) = (d, k) is eligible if i, $E_{dk} < SL$ and $E_{dk} + l_i \leq T$.*

In phase 2, surgical instruments are assigned to the surgeries. Since instruments can be sterilized overnight, at the beginning of the next day, all surgical instruments are considered sterilized and available. Therefore, Phase 2 is repeated for each day separately. For each day, phase 2 begins by creating a list of instruments. The time of availability A_m for each instrument m is set to zero. The time of availability is the earliest time an instrument is available for use. The first surgery is then selected and the *eligible* instruments are assigned to the surgery. If there is no eligible instrument available, a rental instrument is assigned for surgery. The process continues until all surgeries have surgical instruments assigned. For each instrument, we keep track of information on the surgeries to which it is assigned. Once an instrument is used for surgery, it must go through a sterilization process. This is accounted for by adding the duration of sterilization with the duration of surgery. By doing so, we keep track of the time at which it becomes available again. This process is repeated for all days. Once an initial instrument assignment has been generated, the total instrument-related cost such as the usage cost and the rental cost is computed.

Definition 2. *Let S_i be the start time of surgery i on day d. Let l_i and γ be the duration of surgery i and sterilization time, respectively, in time slots. Let D_{ir} be the requirement of surgery i for the type of instrument r. Let A_m be the earliest time the instrument m is available. Let P_{mr} be a binary matrix that denotes whether the instrument m is of type r. Then an instrument m is an eligible instrument for surgery i, if $A_m < S_i$ and $P_{mr} = 1$.*

Definition 3. *The time of availability A_m is updated as $A_m = S_i + l_i + \gamma$*

The pseudocode of phase 1 of the heuristic H2 is given below. The algorithm begins by initializing two empty sets, set $\mathbb{N}$ surgeries which is a randomly ordered set. The set β is a set of all open (Day, OR) tuple. We maintain a list of surgeries assigned to each (day, OR) tuple denoted by β_{dk}. We begin by assigning the first surgery to the first (Day, OR) tuple in the set β. For subsequent surgeries, we select the first eligible (Day,OR) tuple and assign the surgery. Once the surgical assignments to (Day, OR) tuples have been generated, we calculate the cost of the schedule given by Z_1. We keep track of the incumbent cost of the schedules, Z_S. The number of available (Day,OR) tuples is then reduced until the problem becomes infeasible. The minimum cost solution is selected as the starting solution for phase 2.

Algorithm 4 Pseudocode of Phase 1 of H2

procedure ▷
    Step 0: Initialize sets $\mathbb{N}$, β, $Z_1 = \infty$ and $\beta_{dk} = []$
    Step 1:
    for $i \in L$ **do**
        for $(d,k) \in \beta$ **do**:
            if (d,k) **is** eligible :
                $\beta_{dk} \cup i$
            else next (d,k)
        end for
    end for
    Step 2: Sequence surgeries in the order they were assigned
    Step 3: Compute Z_1
    Step 4:
    if $Z_S < Z_1$
        $Z_1 = Z_S$
    Step 5: Update β and go to step 1

The pseudocode of phase 2 of the heuristic H2 is given below. The algorithm begins by initializing β_d which is a list of surgeries assigned to day d, an instrument assignment matrix V_i, a time of availability matrix A for each instrument m. For each day, surgery i with the earliest start time is chosen from β_d. Next, the instrument requirement for surgery i is checked. If an eligible instrument is available for use, the instrument is assigned to the surgery i, otherwise a rental instrument is assigned. Once an instrument m is assigned for surgery i, the time of availability matrix for that instrument is updated. This process is repeated for each day in the planning horizon. The total instrument-related cost of the solution is denoted by Z_I.

Algorithm 5 Pseudocode of Phase 2 of H2

```
procedure                                                    ▷
    Step 0: Initialize sets ℕ, β, Z_1 = ∞ and β_d
    Step 1:
    for d ∈ δ do
        for i ∈ β_d do:
            if instrument m is eligible:
                V_i ∪ m
            else V_i ∪ R
        end for
    end for
    Step 3: Compute Z_I
```

This completes our initial solution heuristic H2. The total cost of the initial schedule is given by $Z_F = Z_S + Z_I$. The resulting solution is used as a starting point for our Tabu search.

3.3.2 Neighborhood definition

In scheduling problems, a solution is represented by a series of surgeries. A neighborhood for a given solution is defined as the set of permutations that can be created by a certain perturbation of the current solution. Some common perturbation schemes used for scheduling problems are adjacent exchange, random exchange, and insertion [47]. In the proposed TS algorithm, we use three perturbation methods that are based on two exchange schemes: 1) two exchanges in the same OR (2EX1), 2) two exchanges between different ORs of the same day (2EX2), and 3) two exchanges between the two days in any OR (2EX3).

A solution to our problem is represented by the sequence of surgeries that will be carried out in each (Day,OR) tuple. The following is a sample representation of an OR schedule.

Day 1, OR 1: Surgery 1 - Surgery 2 - Surgery 3 - Surgery 4

Day 1, OR 2: Surgery 5 - Surgery 6 - Surgery 7

Day 2, OR 1: Surgery 8 - Surgery 9 - Surgery 10 - Surgery 11

Day 2, OR 2: Surgery 12 - Surgery 13 - Surgery 14 - Surgery 15

Example 1. *Let* $\pi_{d_d k_k}$ *be the current sequence of surgeries in (Day d, OR k). Let* $\pi_{d_1 k_1} = [1,2,3,4]$, $\pi_{d_1 k_2} = [5,6,7]$, $\pi_{d_2 k_1} = [8,9,10,11]$, *and* $\pi_{d_2 k_2} = [12,13,14,15]$.

We will use Example 1 to formally define the exchange schemes.

Definition 4. *Let $\pi_{d_1k_1}$ be the current sequence of surgeries in (Day 1, OR 1). Then, a neighbor of $\pi_{d_1k_1}$ using the 2EX1 scheme is obtained exchanging surgeries at positions i, j. Surgery $\pi_{d_1k_1}[i]$ is then swapped with surgery $\pi_{d_1k_1}[j]$.*

In Example 1, $\pi_{d_1k_1} = [1,2,3,4]$. Let i = 2, j = 3, then the neighbor according to 2EX1 would be $\bar{\pi}_{d_1k_1} = [1,3,2,4]$.

Definition 5. *Let $\pi_{d_1k_1}$ and $\pi_{d_1k_2}$ be current sequence of surgeries in (day 1, OR 1) and (day 1, OR 2). Then, a neighboring solution is obtained using the 2EX2 scheme by swapping the surgeries in position i from $\pi_{d_1k_1}$ and in position j from $\pi_{d_1k_2}$. Surgery $\pi_{d_1k_1}[i]$ is swapped with surgery $\pi_{d_1k_2}[j]$.*

In Example 1, $\pi_{d_1k_1} = [1,2,3,4]$ and $\pi_{d_1k_2} = [5,6,7]$. Let i = 2 and j = 3, then the neighbor according to 2EX2 would be $\bar{\pi}_{d_1k_1} = [1,7,3,4]$ and $\bar{\pi}_{d_1k_2} = [5,6,2]$.

Definition 6. *Let $\pi_{d_1k_1}$ and $\pi_{d_2k_1}$ be the current sequence of surgeries in (day 1, OR 1) and (day 2, OR 1). Then, a neighboring solution using the 2EX3 scheme is obtained by exchanging the surgeries in position i from $\pi_{d_1k_1}$ and position j from $\pi_{d_2k_1}$. Surgery $\pi_{d_1k_1}[i]$ is swapped with surgery $\pi_{d_2k_1}[j]$.*

In Example 1, $\pi_{d_1k_1} = [1,2,3,4]$ and $\pi_{d_2k_1} = [8,9,10,11]$. Let i = 1 and j = 1, then the neighbor according to 2EX3 would be $\bar{\pi}_{d_1k_1} = [8,2,3,4]$ and $\bar{\pi}_{d_2k_1} = [1,9,10,11]$.

To speed up the search process, we only consider feasible moves in a neighborhood. This helps us to eliminate several solutions from consideration. Before any move is considered, a feasibility check is performed. The definition of a feasible neighboring move for our neighborhood is given below.

Definition 7. *Let E_{k_1} and E_{k_2} be the end times of the last surgery in the ORs k_1 and k_2, respectively. Let surgery $i \in k_1$ and has duration l_i and surgery $j \in k_2$ has duration l_j time slots. A swap (i, j) is feasible if $E_{k_1} - l_i + l_j \leq T$, and $E_{k_2} - l_j + l_i \leq T$.*

Candidate Lists

Examining the entire neighborhood with TS provides high-quality solutions, in general. However, doing so is computationally expensive [47, 66]. For this reason, it is important to apply TS in conjunction with a strategy that can isolate regions of the neighborhood that have the desired qualities. One of the prominent strategies is called *Preferred attribute candidate lists*. With this strategy, certain attributes of the moves that are expected to be attributes of good solutions are put on a list of candidates [66]. It can be advantageous to isolate certain attributes of moves that are expected also to be attributes of potentially good solutions, and to limit consideration to those moves whose composition includes some portion of these *preferred* attributes. Some surgeries are generally better candidates than others to be scheduled early or later in the sequence. The candidate list considers swaps whose composition includes at least one of these preferred attributes. For example, in the 2EX2 neighborhood, moves that minimize overtime in one OR and idletime in another OR are potentially good moves to evaluate. These moves are added to our *Preferred attribute candidate lists* for the 2EX2 neighborhood. Moves on this list are evaluated first to perform a modified search within each neighborhood, followed by an evaluation of other moves from the neighborhood. Next, we describe the modified neighborhood search methods for each neighborhood.

Preferred attribute move for 2EX1: In the first neighborhood 2EX1, swapping two surgeries in the same OR does not affect overtime or idletime. In that case, we search for two adjacent surgeries that have a common instrument requirement. Examining the moves in which adjacent surgeries do not have a common instrument requirement is likely to yield improved solutions. For example, suppose that adjacent surgeries positioned at i and $i+1$ have a common instrument requirement. If surgery at position i is swapped with surgery $i-1$ or surgery at position $i+1$ is swapped with $i+2$, then it is likely that surgeries previously positioned at i and $i+1$ can reuse the same instrument after sterilization. As a result, moves $(i,i-1)$, $(i+1,i+2)$ are added to the candidate list. This is done for each (Day,OR) tuple. If there are no moves on the candidate list, we continue with the definition 4 of the 2EX1 neighborhood where the values of i and j are generated randomly.

Preferred attribute move for 2EX2: In the second neighborhood 2EX2, we swap surgeries in two different ORs the same day. In this case, swapping two surgeries can affect overtime or idletime. Since OR-related costs are very high, examining the moves that minimize OR-related costs is likely to yield improved solutions. For example, consider a schedule with cost $Z_{s^{old}}$ suppose that k_1 and k_2 are the two ORs in the neighborhood under consideration on day d and $E_{dk_1} < SL$ and $E_{dk_2} > SL$. This indicates that idletime and overtime costs are incurred in ORs k_1 and k_2, respectively. For each swap between surgery $i \in \beta_{dk_1}$ with duration l_i and surgery $j \in \beta_{dk_2}$ with duration l_2, we calculate new values of $E_{dk_1} = E_{dk_1} - l_i + l_j$ and $E_{dk_2} = E_{dk_2} - l_j + l_i$. With the new ending times for each OR, we calculate the new cost of the schedule $Z_{s^{new}}$. If $Z_s^{new} < Z_s^{old}$, move (i,j) is added to the candidate list. This process is repeated for each day on the planning horizon as swapping two surgeries on the same day does not affect the OR-related costs on other days. If there are no moves in the candidate list, we continue with the definition 5 of the 2EX2 neighborhood where the values of i and j are generated randomly.

The preferred attribute move for 2EX3: In the third neighborhood 2EX3, we swap surgeries in two different (Day, OR) tuples on different days. In the modified search for 2EX3, two different types of moves are added to the candidate list that exibit preferred attributes. The first type of move minimizes OR-related costs in two ORs on different days d_1 and d_2. We search for two tuples of different days for which

$E_{d_1k_1} < SL$ and $E_{d_2k_2} > SL$. Then the exact same process as described above is followed for the two (Day, OR) tuples under consideration. Once moves that minimize the OR-related costs are added to the candidate list, we prioritize moves that can likely reduce the instrument-related costs. For example, let surgery $i \in \beta_{d_1}$ and uses rental instrument of type R_1 and surgery $j \in \beta_{d_2}$ and uses rental instrument of type R_2. We add move (i, j) to the candidate list. This reduces the requirement for the instrument of type R_1 and R_2 by 1 for days d_1 and d_2, respectively. If these surgeries share an instrument requirement, then these moves are added at the end of the list. If there are no moves on the candidate list, we continue with the definition 6 of the 2EX3 neighborhood where the values of i and j are randomly generated.

3.3.3 Move strategy

The move strategy determines how the neighbors are visited. The commonly used move strategies are the best move and the first-better move [47]. In the first-better-move strategy, neighbors are evaluated one after the other. If an improving solution is found, a move is made to that solution. In the best-move strategy, all neighbors are evaluated, and then only a move is made to the most improving solution. Evaluating all neighborhoods is not computationally efficient if the neighborhoods are large. Since the 2EX2 and 2EX3 neighborhoods are relatively large, we use the first-better-move strategy.

The move value, the difference between the objective function value of the current incumbent solution and that of a neighboring solution must be calculated for all neighboring solutions considered. However, in our problem structure, each day is independent of each other and, therefore, the evaluation of a neighbor in 2EX1 and 2EX2 only changes the cost for the corresponding day. The evaluation of the neighbors in 2EX3 changes the costs for two days of consideration. Therefore, for efficiency, only changes in the days affected by the move are included in that calculation, significantly reducing the effort required for the evaluation of the move.

3.3.4 Tabu List

The Tabu list T_l stores all forbidden moves called *Tabu moves*. The size of the list T_l is bound by a parameter L_t, called the *Tabu list size*. When the cardinality of T_l is L_t, before adding a new element to T_l, the oldest element is removed from the list. The size L_t of the Tabu list could be fixed or could be dynamically changed through certain adjustments [64]. Via an initial set of experiments, we establish that a Tabu list of size 6 is appropriate for our experiments and implement it as a first-in, first-out queue. Our definition of a Tabu move is given below.

Definition 8. *Let i and j be the two surgeries swapped in iteration n. Then swapping of surgeries (i, j) or (j, i) is considered Tabu for the next N_t iterations.*

In example 1, $\pi_{d_1k_1} = [1,2,3,4]$. Let i = 2, j = 3, then the neighbor according to 2EX1 would be $\bar{\pi}_{d_1k_1} = [1,3,2,4]$. We add a tuple (3,2) to the Tabu list. For the next N_t iterations, any swap between these surgeries will not be considered.

3.3.5 Stopping condition

The algorithm begins with the 2EX1 neighborhood for each day and each OR and searches for the solution space following the rules described above. As the 2EX1 neighborhood is comparatively small, we explore all the neighbors in the 2EX1 neighborhood. After that, the algorithm explores the 2EX2 neighborhood of the current solution. This neighborhood is explored for each day for a certain number of iterations. If no better solution is found, the algorithm switches to the 2EX3 neighborhood. We call this the intensification phase, or inner loop. In this phase, we search the neighborhood of the current solution and move to the best feasible neighbor solution that is not Tabu. The inner loop is terminated when a prespecified number of iterations is reached.

In the diversification phase, that is, the outer loop, there are two main diversification techniques used in the literature, restart diversification and continuous diversification. In the continuous diversification process, diversification measures are included in regular neighborhood search procedures. In restart diversification, the search is moved towards unvisited regions by introducing few rarely used components to the current solution. The TS algorithm is restarted from that point on. We use restart diversification where we restart the TS algorithm with a different initial solution. To begin with a new initial solution, we randomly order the set $\mathbb{N}$ and run the heuristic H2 again. When this is done, the surgeries are assigned to different (Day,OR) tuples, and therefore different solution spaces are searched. The outer loop is terminated when a pre-specified number of iterations is reached.

3.3.6 Description of the TS algorithm

The pseudocode of the algorithm presented in 6 combines the aforementioned elements of the Tabu search. The algorithm starts by initializing some parameters first. The parameters N^I and N^o denote the number of inner and outer iterations to be performed. The Tabu list, T_l, is initially an empty set. For each outer iteration, we start with a new current solution X_c. For each inner iteration, the 2EX1 neighborhood is explored in search of a better solution. If an improving non-Tabu solution X_b is found, we set $X_c = X_b$. The Tabu list is updated after we change the current solution. We increase the inner iteration number and continue with the same neighborhood. If an improving solution is not found, the neighborhood is changed to 2EX2 and then to 2EX3. Both 2EX2 and 2EX3 are evaluated for a certain number of moves in addition to all the moves on the preferred attributes candidate list. At the end of an iteration, if an improving neighboring solution is not found, we select the neighbor with the least non-improving cost and call it neighbor X_b, and set it to current solution, i.e., $X_c = X_b$. After N^I iterations, we diversify the search by restarting with a new initial solution.

Algorithm 6 Pseudo code for TS algorithm

procedure ▷
  Initialize N_{in}, N_{out}, T_l = []
  while $N_{out} < N^o$ **do**
    Obtain initial solution X_c,
    while $N_{in} < N^I$ **do**
      Perform search in 2EX neighborhoods
      Select first most improving/ least non-improving, non Tabu neighbor X_b
      Set $X_c = X_b$, $Z_c = Z_b$
      Update T_l
      if $Z_b < Z^*$
        $Z^* < Z_b$
      $N_{in} = N_{in} + 1$
    end while
    $N_{out} = N_{out} + 1$
  end while

3.4 Computational Results

The objective of our computational analysis is to evaluate the performance of the TS algorithm in terms of execution time and the quality of the solutions provided. We demonstrate this using the same case studied in Chapter 2. The same set of surgical cases from Chapter 2 was used to solve the integrated OR scheduling problem using TS. To convert the duration of surgeries into time slots l, we round the actual duration to the nearest half-hour. We solve the problem of creating a weekly OR schedule, and hence the number of days δ is chosen to be 5. Our data analysis indicates that different services do not share more than 3 ORs. Therefore, the number of ORs, K, is set to 3 ORs. The number of surgeries to be scheduled is set to $N = 30$. This number is also inspired by data that indicate that up to 2-3 surgeries are scheduled in an OR on a typical day. We consider instruments of six different types, that is, $R = 6$, and the inventory level, I_r, is set to two instruments per type. The requirement matrix for surgical instruments was generated randomly for each type of surgery. We assume that surgeries may not require more than 1 instrument of each type. From the data collected from GMH, we observed that instances of multiple expensive instruments of the same type are used for one surgery are rare.

We set the following problem parameters: fixed cost of opening the OR, c^k = \$360, session length $s = 16$, total number of time slots $T = 20$, overtime cost, c^o =\$100, and the idle time cost, c^u = \$22.5. We assume that using 2 hours of overtime is equivalent to opening a new OR. We set the usage cost as $c^F = \$18$ and the instrument rental cost as $c^R =$ \$36. The opening cost of the OR, the overtime cost, and the idle time costs are kept the same for all experiments. The sterilization time, γ, is set at 6 time intervals, that is, 3 hours, based on an expert opinion from GMH. We assume that the instruments are sterilized immediately after surgery. As the sterilization department works for 24 hours in shifts, surgical instruments can be sterilized overnight and it is assumed that they are available for use at the start of the next day.

By changing the parameters of the TS algorithm, we conducted an initial set of experiments. In these experiments, we tested different values of inner and outer iterations, different sizes of Tabu list, different limits on how many swaps to be evaluated before changing the neighborhoods, etc. We observed that Tabu list of size 6 works best for our problem instances. We run the algorithm for 90 minutes to complete a sufficient number of iterations of small, medium, and large-sized instances. The number of inner iterations is set to 30 iterations. In each inner iteration, we create the preferred attribute lists. All neighbors in this attribute list are examined first. Additionally, we examine 50 and 75 additional neighbors in neighborhoods 2EX2 and 2EX3, respectively. After all inner iterations are performed, we begin with a new solution. We perform 50 outer iterations to examine a diverse set of solutions. The TS algorithm is coded using the Python programming language and is executed on a high performance computer cluster with an HP computer with an Intel Xeon processor with 12 cores and 42GB of memory.

We distinguish between small, (i.e., $N = 10$), medium, (i.e., $N = 20$), and large, (i.e., $N = 30$) problem instances. Each set of problems consists of 10 different instances and is solved using the LD-SH method described in Chapter 2. When problem (P) is solved using LD-SH, the lower bound LB^{max} is provided by LD, while the best upper bound UB is provided by the SH counterpart of LD-SH. We solve these problem sets again using the TS algorithm. When using the TS algorithm, the least cost solution is used as the upper bound UB, as this is the best value of the objective function. The lower bound used for the comparison is the same as that of LD. The error gap for these experiments is calculated as follows.

$$\text{Error Gap}(\%) = \frac{UB - LB}{UB} * 100 \tag{3.1}$$

Tables 3.1, 3.2, and 3.3 compare the results of TS for mixed-duration problem instances with the LD-SH developed in Chapter 2. We also report the minimum, maximum, and average running time at which the incumbent solution found via TS was recorded.

LD			TS			
	Error Gap %	Run Time (Hours)	Error Gap %	Run Time (Hours)	Incumbent Sol. Time (Hours)
Avg	0.00	0.01	0.00	1.50	0.006
Min	0.00	0.01	0.00	1.50	0.00
Max	0.08	0.02	0.00	1.50	0.06

Table 3.1: Results for Computational Performance (Small)

	LD		TS		
	Error Gap %	Run Time (Hours)	Error Gap %	Run Time (Hours)	Incumbent Sol. Time (Hours)
Avg	1.60	1.10	0.60	1.50	0.03
Min	0.00	0.10	0.00	1.50	0.00
Max	8.07	5.00	8.07	1.50	0.21

Table 3.2: Results for Computational Performance (Medium)

	LD		TS		
	Error Gap %	Run Time (Hours)	Error Gap %	Run Time (Hours)	Incumbent Sol. Time (Hours)
Avg	5.61	3.90	3.77	1.50	0.05
Min	0.34	0.37	0.00	1.50	0.00
Max	10.22	5.00	10.22	1.50	0.11

Table 3.3: Results for Computational Performance (Large)

Observation 1: For small-sized problems, TS algorithm finds an optimal solution for every instance. For medium- and large-sized problems, the average error gap is 0.6% and 3.77%, respectively. This indicates an improvement in solution quality on average. However, the maximum error gap is 8.07% and 10.22%, respectively, which is the same as the LD-SH method. This indicates that some problems were inherently difficult to solve. LBs from LD-SH not being tight can also be one of the factors for such a large error gap in some of the instances. In few instances, the TS solution finds the optimal solution for large-sized problems which LD-SH fails to do.

Observation 2: For small-sized problems, TS algorithm finds the optimal solution for each instance. Although we did not stop the TS algorithm for 1.5 hours, the average time to find the incumbent solution was 0.006 hours. This indicates that most of the optimal solutions were obtained almost instantaneously. For medium-sized problems, the average incumbent solution time was 0.03 hours, much shorter than that of the LD-SH algorithm. For large-sized problems, average time to find the best solution was 0.05 hours. This indicates that for medium- and large-sized problems, the TS algorithm did not improve after finding the solution in earlier iterations

Quality of Upper Bounds

To understand the quality of the solutions provided by the TS algorithm, we compare the best value of the objective function of TS, Z^t, with the best upper bound values, UB^{min}, reported in Chapter 2. Similar to Chapter 2, we distinguish between short, $l < 4$ (hours) and long $l > 4$ (hours) surgeries. These experiments also allow us to examine under what conditions the TS algorithm performs better compared to the LD-SH method. In this part of our computational study, we solve all the problems described in Table 2.5 in Chapter 2. Tables 3.4, 3.5, and 3.6 summarize the results of these experiments. We define the UB gap as the percent improvement in total costs if we use the TS algorithm instead of the MIP based method. Negative values indicate that the TS algorithm obtained an improved feasible solution compared to the solutions obtained using LD-SH. The UB gap for these experiments is calculated as follows.

$$\text{UB Gap}(\%) = \frac{Z^t - UB^{min}}{Z^t} * 100 \tag{3.2}$$

	UB Gaps %		
Problems	Minimum	Average	Maximum
Problem 1	0.00	0.00	0.00
Problem 2	-0.87	-0.17	0.74
Problem 3	-1.19	-0.13	5.94

Table 3.4: Result of UB comparison for short-duration surgeries

	UB Gaps %		
Problems	Minimum	Average	Maximum
Problem 4	-0.81	-0.08	0.00
Problem 5	-0.62	-0.18	0.00
Problem 6	-0.80	-0.33	4.91

Table 3.5: Result of the UB comparison for the mix of short and long-duration surgeries

	UB Gaps %		
Problems	Minimum	Average	Maximum
Problem 7	-0.94	-0.19	1.46
Problem 8	-0.91	0.15	2.99
Problem 9	-0.68	1.43	5.88

Table 3.6: Result of the UB comparison for long-duration surgeries

Observation: For short, mixed, and long-duration surgeries, TS provides high-quality solutions for all problem sizes. For large-sized problems 3, 6, and 9, TS, on average, finds a solution within -0.13%, -0.3%, and 1.43% of the LD-SH method, respectively. For problems 1 to 7, the average error gap is less than or equal to 0%. This indicates that by using the TS algorithm, the feasible solution can be improved. TS performs excellently for small and medium-sized instances. For large-sized problems, the performance of TS deteriorates. Although the average UB gap is less than 1.43%, the maximum error gaps are 5.94%, 4.91%, and 5.88% respectively. Large problem instances have a larger neighborhood to explore, and we stop the search after 1.5 hours have elapsed. These could be some of the factors due to which the UB gaps are significantly large in some instances. TS has a worse performance for long-duration surgeries in general compared to LD-SH.

As TS chooses the neighboring solution first via a candidate list of neighbors and then via regular neighborhood search, the trajectory of solutions examined is different from that of Lagrangean decomposition-based heuristics. In the previous chapter, we conducted sensitivity analyses with respect to cost ratios (see Figure 2.4). We conduct the same sensitivity analyses with respect to f/v ratios using TS algorithm. For these experiments, we vary the rental instrument cost c^R from 18, 36, 90, 180, 360, 450 to generate ratio values ($r_1 - r_6$) as 0.05, 0.1, 0.25, 0.5, 1, 1.25 respectively. The results of the TS experiments were compared with those of the LD-SH and LD-BD methods. Figure 3.1 summarizes these results. The blue, black, and red lines indicate the error gap for the LD-SH, LD-BD, and TS heuristics. The results indicate that for all values of c^R, the performance of the TS algorithm is comparable to LD-SH. For higher values of c^R, TS provides excellent results in general.

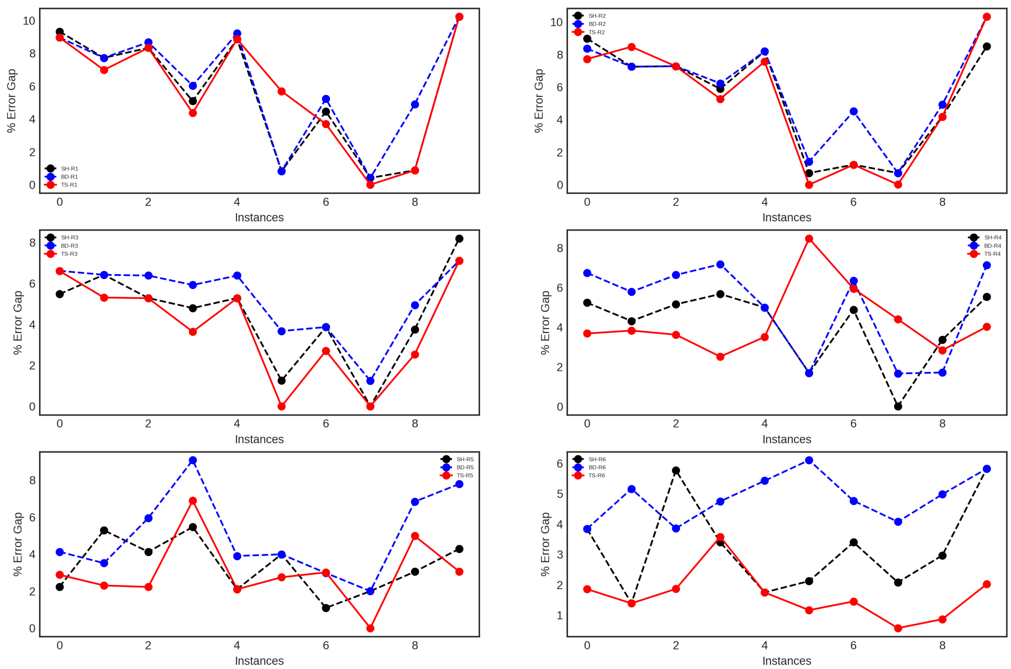


Figure 3.1: Sensitivity analysis w.r.t. Cost Ratios

3.5 Conclusion

The research and models presented here are motivated by the opportunities for improvement observed in GMH's inventory management and OR scheduling practices. The Lagrangean decomposition algorithm developed in the previous chapter takes several hours on average to find high-quality solutions. Furthermore, each iteration of the algorithm sometimes takes several minutes. This indicates that there is an opportunity to improve these solutions. We faced challenges when trying to solve larger-sized problems using the methods developed in the previous chapter and commercially available solvers. Moreover, most hospitals and clinics may not have access to the state-of-the-art solvers. In this chapter, we propose a new construction heuristic H2 and a TS algorithm to solve an integrated OR scheduling problem to minimize the cost of opening the ORs, overtime, and idle-time along with the costs of using and renting instruments.

Heuristic H2 obtains the initial OR schedule with instrument assignment in less than 1 second on average for the large-sized problems. This is a significant improvement over the heuristic H1, which requires one to spend several minutes to solve two MIP models (Models S and F) to obtain the initial solution.

In our TS algorithm, we use three neighborhoods. Neighborhoods are changed after a certain number of moves have been evaluated. After all three neighborhoods have been examined, we begin with a new initial solution. A candidate list of solutions with the preferred attributes is created separately for each neighborhood to prioritize the examination of favorable solutions. The candidate list and the changing neighborhood structure together prevent examining the entire neighborhood. The results of TS algorithm are compared with the solutions obtained via the Lagrangean decomposition-based algorithm. The running time of 1.5 hours is set for all problems solved with the TS algorithm. For small-sized problems, the TS algorithm finds the optimal solution for every instance. For medium and large-sized problems, we obtain solutions within the error gaps of 0.5% and 4%, respectively. We also report the time in which the incumbent solutions were found. For all instances, the average time to the incumbent solution is significantly shorter than the average running time of LD-SH algorithms. In some instances, the best solution does not improve for a long time. In general, TS provides high-quality upper bounds compared to LD-SH. Although the average UB gap is less than 1.43%, UB gaps can be as large as 6%. The performance of the TS algorithm worsens when only long-duration surgeries are considered. These experiments indicate that the TS algorithm yields solutions that are better than those obtained via LD-SH in a significantly shorter period of time. These solutions are found using the 40% of time used by LD-SH on average.

Chapter 4

Conclusion and Future Research

In this dissertation, we study tactical and operational problems that arise in the perioperative services departments of hospitals. We focus on scheduling operating rooms (OR) and managing necessary resources such as equipment, and surgical instruments. Our work demonstrates how the use of mathematical and simulation optimization methods in novel ways can improve healthcare systems. These methods have been used for healthcare operations management for decades. Our contribution to the existing literature lies in the study of newly developed models and solution methods in the areas of material handling, OR scheduling, and inventory management of surgical instruments in hospitals. For these studies, we collaborated with Greenville Memorial Hospital (GMH) in South Carolina.

Our first study was motivated by the inefficiencies observed in the GMH material handling system. GMH staff reported long lines of AGVs waiting for the elevator on the mezzanine floor after material handling activities started in the afternoon. Congestion caused by AGVs contributes to delays in delivering the required surgical material, including surgical instruments, to the ORs. Congestion also affects the delivery of soiled surgical instruments, further extending sterilization activities. Delays due to the material handling activities impact the utilization of AGVs, surgical instruments, sterilization equipment, and personnel time. These delays also force GMH to use the rental and consigned instruments. The simulation-optimization framework presented in Chapter 1 enables hospitals to identify the factors that affect the performance of material handling and inventory management systems, and develop solutions that improve their efficiency. Our work highlights the role of coordinating decisions between material handling and inventory management to improve the level of service provided by ORs. We find that using fewer AGVs leads to reduced congestion, as evidenced by the average travel time and the corresponding range of travel time, which is narrower. On the contrary, when the objective is to minimize the task completion time, the simulation experiments suggest using relatively more AGVs. The solution recommended by the simulation-optimization method uses fewer AGVs on all days than the current policy of using 11 AGVs on each day. Upon the implementation of this solution, it was found that the task completion time is not different from that of current practice. However, the recommended solution takes on average 2.5 fewer minutes to deliver each surgical case cart than the current practice. To validate these results, we conducted a pilot study at GMH. The results of this pilot study indicated that travel time was significantly reduced with our recommended solution. Our simulation experiments also indicate that the current material handling approach is more sensitive to the level of surgical instrument inventory, whereas the two alternative approaches that we propose are not. When there is a low inventory level, the two-batch and JIT approaches can reduce surgical delays by 83% and 95%, respectively. By choosing alternative approaches, GMH could reduce its inventory by 2-4 units for every specialty. Both of these approaches also use fewer AGVs than the current approach. The staff at our partner hospital have considered the models we developed and the experimental results as valuable inputs, and have implemented our recommendations in some capacity.

This study can be extended in multiple directions. First, we focus only on the material handling and inventory management of surgical instruments. In addition to these, there are several other material movements that occur simultaneously in the hospital. During the pilot study, we learned that GMH prioritizes food delivery over other material movements. One way to extend this study is to optimize the material handling of all delivery services, such as food, linen, surgical instruments, and trash, simultaneously. The interaction of other delivery services with the movement of surgical instruments and its impact on the required inventory levels could be investigated. Furthermore, in our models, we do not have disruptive events such as patient no-shows, cancellations, incorrect case cart loading, AGV failures, etc. Investigating the effects of one or a combination of these factors can be a part of future work. Furthermore, an important assumption for this study is that each instrument is being used once it has been moved to the OR and, therefore, sterilized. Several studies indicate that instruments may not be opened despite being on the doctor's preference card (DPC). An extension of this study is to develop a stochastic tray optimization model where there is a certain probability for each surgical instrument to be actually used in surgery.

Our second study shows the importance of coordinating OR scheduling and inventory management decisions in improving OR efficiency and reducing costs. In particular, this work shows that integrating these decisions can increase the utilization of ORs and surgical instruments while reducing the cost of the system. We propose easy-to-implement solution methods, including a construction heuristic and a Lagrangian decomposition-based heuristic, and evaluate them. These solution approaches outperform the commercial solver, Gurobi, in terms of running time and solution quality. The use of the integrated OR scheduling model results in 1 to 4.3 fewer instruments being rented each week on average for all surgery durations considered compared to current practice. This corresponds to a 1-3% cost reduction per week. The greatest improvement is observed when scheduling short-duration surgeries. This is because many short-duration surgeries are scheduled in a day. Thus, instruments can be reused frequently. We find that the number of ORs and the length of planning horizon do not impact overtime and idle time costs as much. However, the total available (day,OR) time has a high impact on OR-related costs. The number of instruments used, reused, and rented is affected by both the planning horizon and the number of ORs available each day. The planning horizon has a greater impact than the number of ORs on these factors. The algorithms we developed to solve the integrated OR scheduling problem show significant improvement in terms of solution quality and running time when compared to directly solving the original problem formulation with the Gurobi optimization solver. On average, the proposed algorithms improve optimality gap by 0.3% for small-sized problems and 5.5% for medium and large-sized problems. The proposed algorithms reduce running time by 95%, 56%, and 15% on average for small, medium, and large-sized problems to find these improved solutions.

The models presented here are particularly suitable for hospitals using an open scheduling strategy; and other health-care facilities can also learn from these practices. One of the possible extensions to our model is to study the stochastic version of the integrated OR scheduling problem. Certain parameters in our formulation are usually unknown at the time the schedules are generated; however, their values when they are realized can significantly impact costs. For example, while some surgeries are completed on time or ahead of schedule, while others take longer than expected. Unavailable surgical materials or some complications can also delay surgeries. These stochastic surgery durations can affect overtime, idle time, and instrument availability. The algorithms presented in this study, for example, Bender's decomposition algorithm (BDA), can be extended to solve a stochastic version of our problem. However, the run-time performance of such a stochastic program needs to be investigated and potentially improved. The cuts developed for BDA are also valid for the stochastic integrated OR scheduling problem. However, feasibility cuts will have to be added if a feasible assignment of surgeries to the tuple (day,OR) is not found. We currently do not include patient no-shows, cancellations, and emergency surgeries in our formulation. Studying the effects of one or a combination of these factors and developing models and methods for extended models can become part of future work in this area.

Finally, our third study develops solution approaches to find near-optimal solutions to the integrated OR scheduling problem without the use of state-of-the-art optimization solvers. We propose easy-to-implement solution methods that include a new construction heuristic and a Tabu Search (TS) algorithm, and evaluate them. To improve the running time of TS, we implement a changing neighborhood structure to search for a solution in different regions of the feasible space. To further improve the run time, we implement a preferred attribute list for the neighborhood search. Solutions with certain favorable attributes are placed on a candidate list and examined first before the rest of the solutions. We evaluate our metaheuristic approach via a case study using data from GMH. These solution approaches outperform the commercial solver, Gurobi, and previously developed Lagrangian heuristic-based approaches in terms of running time and solution quality. For small-sized instances, TS finds the optimal solution for every instance. For medium-sized instances, TS finds solutions within the 0.5% error gap. For large-sized instances, TS finds solutions within a 4% error gap. TS algorithm reduces the running time by 60% on average to find an improved solution compared to LD-SH. This shows that hospitals can use this method to determine their schedule for each week.

Meta-heuristics such as TS have shown their efficiency in solving various complex optimization problems. They generate a lot of data in the search process, including the sequence of solutions, moves, local optima, and bad solutions. Machine learning can be used to analyze the data to extract useful information that can be used to improve search performance. Algorithms have been developed to improve metaheuristics in terms of convergence speed, solution quality, and robustness. Thus, one possible extension of this study is to consider these algorithms and compare the results with the proposed TS method.

Appendices

Appendix A Appendix Chapter 1

Table 1: Model Inputs and Outputs

Component	Details
Model Output (Responses)	
Response 1	Average and standard deviation of trip time of AGVs carrying clean surgical case carts.
Response 2	Average and standard deviation of trip time of AGVs carrying soiled surgical case carts.
Response 3	Average completion time for all surgical case carts.
Model Input (Experimental Factors)	
Input 1	The path of AGVs.
Input 2	Number of AGVs required (maximum up to 11 AGVs).

Day	Distribution
Mon.	DISC(0.004,0,0.05,30,0.096,60,0.139,90,0.181,120,0.248,150,0.309,180,0.355,210,0.397,240,0.444,270, 0.516,300,0.562,330,0.614,360,0.662,390,0.704,420,0.73,450,0.758,480,0.789,510,0.828,540,0.861,570, 0.889,600,0.895,630,0.917,660,0.928,690,0.932,720,0.946,750,0.952,780,0.954,810,0.959,840,0.961,870, 0.965,900,0.969,930,0.969,960,0.972,990,0.976,1020,0.976,1050,0.978,1080,0.98,1110,0.983,1140, 0.987,1170,0.987,1200,0.989,1230,0.993,1260,0.993,1290,0.998,1320,0.998,1350,0.998,1380,1,1410)
Tue.	DISC(0.006,0,0.035,30,0.088,60,0.146,90,0.203,120,0.259,150,0.307,180,0.359,210,0.42,240,0.461,270, 0.518,300,0.572,330,0.61,360,0.653,390,0.689,420,0.72,450,0.752,480,0.789,510,0.821,540,0.85,570, 0.875,600,0.902,630,0.908,660,0.919,690,0.929,720,0.935,750,0.946,780,0.948,810,0.952,840,0.958,870, 0.96,900,0.965,930,0.969,960,0.969,990,0.973,1020,0.975,1050,0.975,1080,0.975,1110,0.979,1140, 0.985,1170,0.987,1200,0.988,1230,0.99,1260,0.992,1290,0.994,1320,0.996,1350,1,1380,1,1410)
Wed.	DISC(0,0,0.004,30,0.033,60,0.075,90,0.133,120,0.18,150,0.264,180,0.308,210,0.353,240,0.399,270, 0.472,300,0.523,330,0.577,360,0.621,390,0.645,420,0.694,450,0.74,480,0.785,510,0.818,540,0.843,570, 0.893,600,0.914,630,0.923,660,0.934,690,0.944,720,0.951,750,0.958,780,0.964,810,0.964,840,0.964,870, 0.969,900,0.974,930,0.976,960,0.978,990,0.978,1020,0.984,1050,0.985,1080,0.985,1110,0.989,1140, 0.991,1170,0.995,1200,0.995,1230,0.995,1260,0.995,1290,0.995,1320,0.995,1350,0.998,1380,1,1410)
Thu.	DISC(0.006,0,0.043,30,0.111,60,0.159,90,0.203,120,0.263,150,0.31,180,0.355,210,0.413,240,0.462,270, 0.513,300,0.544,330,0.592,360,0.638,390,0.675,420,0.713,450,0.754,480,0.787,510,0.818,540,0.838,570, 0.874,600,0.895,630,0.912,660,0.925,690,0.933,720,0.939,750,0.946,780,0.952,810,0.953,840,0.956,870, 0.959,900,0.963,930,0.969,960,0.97,990,0.97,1020,0.973,1050,0.976,1080,0.979,1110,0.98,1140, 0.98,1170,0.98,1200,0.982,1230,0.984,1260,0.989,1290,0.993,1320,0.993,1350,0.994,1380,1,1410)
Fri.	DISC(0.016,0,0.053,30,0.105,60,0.156,90,0.2,120,0.268,150,0.353,180,0.411,210,0.451,240,0.486,270, 0.53,300,0.579,330,0.626,360,0.674,390,0.716,420,0.751,450,0.788,480,0.805,510,0.844,540,0.875,570, 0.9,600,0.912,630,0.923,660,0.933,690, 0.94,720,0.951,750,0.954,780,0.954,810,0.956,840,0.96,870, 0.963,900,0.963,930,0.967,960,0.967,990,0.97,1020,0.97,1050,0.974,1080,0.974,1110,0.975,1140,

Component	Include/Exclude	Justification
Entities		
Surgical Cases	Include	Necessary for modeling the system.
Dummy entities	Include	Necessary for modeling empty transporter movements.
Activities (Departments/locations)		
Picking process (MD)	Include	Key influence on completion time
Inspection (CCSA)	Exclude	Assumption: required material is always available and picked
Manual transport to OR (CCSA - OR)	Exclude	Limited impact on material handling process
Surgery (ORD)	Include	Necessary for modeling
Separation of Instruments from carts (CSSD)	Exclude	Limited impact on material handling process
Sterilization of instruments (CSSD)	Exclude	Limited impact on material handling process
Cart wash (CSSD)	Include	Key influence on completion time
Queues		
OR	Exclude	Assumption: required Ors are always available
For AGVs	Include	Key influence on travel time and completion time
For elevator	Include	Key influence on travel time and completion time
Resources		
AGVs	Include	Required to limit the number of transporters in use
OR	Exclude	Assumption: Ors are always available.
Elevators	Include	Required to move material. Number of elevators/capacity impact travel and completion time
Case cart	Include	Required to move material. Number of case cart impact travel and completion time
Surgical staff	Exclude	Assumption: required staff is always available and picked
Picking process staff	Include	Required to pick material. Number of employees impact completion time
CSSD staff	Exclude	Assumption: required staff is always available and picked
Cart washer	Include	Required for cleaning of case carts. Number of cartwasher impact completion time
Transporter		
AGVs	Include	Experimental factor
AGV Network Paths	Include	Experimental factor

Table 3: Model Scope

Table 4: Model Level of Details

Component	Detail	Include/Exclude	Justification
Entities			
Surgical Cases			
Quantity	TRIA (60,68,75): Monday	Include	Limits to number of entities
	TRIA (65,72,76): Tuesday	Include	Limits to number of entities
	TRIA (60,65,72): Wednesday	Include	Limits to number of entities
	TRIA (69,75,80): Thursday	Include	Limits to number of entities
	TRIA (55,62,69): Friday	Include	Limits to number of entities
Arrival pattern	Constant (24 hours)	Include	Generated every 24 hours
Attributes	Type of entity Clean/Soiled/Washed. Needed to identify the route .	Include	Information required for each entity
Routing	Determine which path transporter takes	Include	Routing dependent on entity type
Dummy entities			
Quantity	One for each surgical case	Include	Limits to the number of entity = 1
Arrival pattern	The entity is generated based on the signal	Exclude	Information required for each entity
Attributes	Type of entity Free/To parking. The attribute is needed to identify route to pickup next case cart or to go to parking location	Include	Information required for each entity
Routing	Determine which path transporter takes	Include	Routing dependent on entity type. Afftects the response (Travel time)
Activities (Departments/locations)			
Picking Proces (MD)			
Quantity	One for each surgical case entity	Include	Model each picking process as it has impact on the response completion time
Nature (X in Y out)	-	Exclude	Sub-components/Instruments are not modelled and so no assembly is represented
Cycle time	TRIA(2,3,5) minutes. Distribution obtained via time study	Include	Required for modelling completion time.
Breakdown/repair	-	Exclude	No breakdowns in picking process
Set-up/changeover	-	Exclude	No changeovers
Resources	Case cart = 1, MD Employee = 1	Include	Identify number of case carts and number of staff required for picking process
Shifts	-	Exclude	No work takes place outside of on-shift time.
Routing	-	Exclude	Routing of entity has no effect
Surgery (OR Department)			
Quantity	One for each surgical case entity	Include	Model each surgery process as it has impact on the responses.
Nature (X in Y out)	-	Exclude	Sub-components are not modelled.
Cycle time	Modeled based on the distribution provided in appendix.	Include	Required for modelling. Has significant effect on responses.
Breakdown/repair	-	Exclude	No breakdowns in surgeries.
Set-up/changeover	-	Exclude	No changeovers
Resources	-	Exclude	All ORs are always available.
Shifts	-	Exclude	No work takes place outside of on-shift time.
Routing	-	Exclude	Routing of entity has no effect.
Cart wash (CSSD)			
Quantity	One for each surgical case entity	Include	Model each washing process as it has impact on the responses.
Nature (X in Y out)	-	Exclude	Sub-components are not modelled.
Cycle time	Fixed : 20 minutes cycle	Include	Influence on resource availability
Breakdown/repair	-	Exclude	Breakdowns are rate and thus not modeled.
Set-up/changeover	-	Exclude	No changeovers
Resources	Cart Washer	Include	Identify number of cart washers required for washing process
Shifts	-	Exclude	No work takes place outside of on-shift time.
Routing	-	Exclude	Routing of entity has no effect.

Table 5: Model Level of Details (continued)

Component	Detail	Include/Exclude	Justification
Queues			
For AGVs			
Quantity	3	Include	All AGVs are individual. One for each department except CCSA.
Capacity	MD- 4 AGVs, CSSD - 1 AGV, and 3 - 2nd Floor soiled cart storage area	Include	Experimental factor.
Dwell time	Until availability of AGV	Include	Afftects the response completion time.
Queue discipline	FIFO	Include	Afftects the response completion time.
Breakdown/repair	-	Exclude	Failures are rare and so have little effect on responses.
Routing	Based on distance of AGV. Shortest distance first.	Include	Routing of AGVs defines the key interaction between system components.
For elevator			
Quantity	3	Include	All elevators are individual. One for each elevator.
Capacity	-	Exclude	Capacity for elevator queues is limited via number of AGVs in the departments.
Dwell time	Until availability of elevator	Include	Afftects the responses travel/completion time.
Queue discipline	FIFO	Include	Afftects the responses travel/completion time.
Breakdown/repair	-	Exclude	Failures are rare and so have little effect on responses.
Routing	To next department using routing logic	Exclude	Routing of entities to and from elevator defines the key interaction between system components.
Resources			
AGVs			
Quantity	-	Include	Experimental factor.
Where required	Identify case carts that require transport	Include	Required to allocate work to AGVs
Shifts	-	Exclude	No work takes place outside of on-shift time
Elevator			
Quantity	Elevator J - 2, Elevator G - 2, Elevator K - 1	Include	Because there maybe fewer elevators than requesting AGVs, it is possible for elevator shortages to be a bottleneck affecting the responses
Where required	Identify the AGVs that require transport between the floors	Include	Required to move the elevators between the floors
Shifts	-	Exclude	No work takes place outside of on-shift time
Case cart			
Quantity	110	Include	Because there maybe fewer case carts than surgeries, it is possible for case cart shortages to be a bottleneck affecting the responses
Where required	-	Exclude	The case carts are only located and assigned at MD.
Shifts	-	Exclude	No work takes place outside of on-shift time
Picking process staff			
Quantity	4	Include	Because there are fewer picking process staff than surgeries, it is possible for staff shortages to be a bottleneck affecting the response completion time
Where required	-	Exclude	The picking process staff is located only at MD.
Shifts	-	Exclude	No work takes place outside of on-shift time
Cart washer			
Quantity	3	Include	Because there are fewer cartwashers than surgeries, it is possible for case cart shortages to be a bottleneck affecting the responses
Where required	-	Exclude	The cartwashers are located only at CSSD.
Shifts	-	Exclude	No work takes place outside of on-shift time
Transporter			
AGVs			
Quantity	20	Include	Because there maybe fewer AGVs than surgeries, it is possible for AGV shortages to be a bottleneck affecting the responses
Type	Guided	Include	Has influence on the responses.
AGV straight velocity	200 units distance	Include	Has influence on the responses.
AGV turning factor	0.5	Include	Has influence on the responses.
Acceleration/Deceleration	0.98 per second squared	Include	Has influence on the responses.
Zone control rule	End	Include	Has influence on the responses/ required for safety.
Shift	-	Exclude	No work takes place outside of on-shift time
AGV Network Paths			
Quantity	53 network links, 2 paths	Include	Necessary for modeling. Has influence on the responses.
Type	Department/elevator links: Spur, Other links: Unidirectional	Include	Has influence on the responses.
Routing	-	Include	Experimental factor.

Table 6: Fraction of Delayed Deliveries by Month

	Simulation of Current System					Simulation of Implemented Solution				
Month	**Delayed per day**	**Average**	**Std**	**Min**	**Max**	**Delayed per day**	**Average**	**Std**	**Min**	**Max**
January	0.613	0.037	0.014	0.02	0.055	0.645	0.038	0.013	0.02	0.055
February	2.214	0.059	0.026	0.02	0.149	2.214	0.060	0.027	0.02	0.149
March	0.581	0.042	0.02	0.021	0.083	0.613	0.043	0.023	0.021	0.104
April	3.633	0.071	0.036	0.02	0.168	3.633	0.071	0.036	0.02	0.168
May	3.161	0.069	0.038	0.019	0.165	3.194	0.069	0.039	0.019	0.165
June	1.433	0.044	0.023	0.021	0.141	1.433	0.044	0.023	0.021	0.141
July	3.161	0.073	0.047	0.019	0.178	3.161	0.073	0.048	0.019	0.178
August	3.645	0.079	0.049	0.019	0.196	3.645	0.080	0.048	0.019	0.196

Table 7: Travel Time by Month (Min)

	Simulation of Current System				Simulation of Implemented Solution			
Month	**Average**	**Std**	**Min**	**Max**	**Average**	**Std**	**Min**	**Max**
January	9.513	2.431	2.725	13.106	6.968	1.798	2.724	11.714
February	9.277	2.607	2.733	13.046	6.766	1.91	2.73	11.714
March	9.404	2.451	2.757	13.106	6.934	1.883	2.721	12.083
April	8.981	2.803	2.72	13.106	6.58	2.01	2.732	11.714
May	9.175	2.685	2.731	13.106	6.706	1.963	2.724	11.714
June	9.234	2.588	2.727	13.106	6.73	1.923	2.721	11.714
July	9.157	2.701	2.746	13.106	6.692	1.933	2.721	11.714
August	9.096	2.743	2.73	13.106	6.727	2	2.72	11.714

Material Handling Process: The research team collaborated with the Perioperative Services Department (PSD) of Greenville Memorial Hospital (GMH). The PSD consists of three departments: the Materials Department (MD) and the Central Sterile Storage Department (CSSD), both of which are located on the mezzanine floor (see Figure **??**), and the Operating Room Department (ORD) located on the second floor. GMH has 46 AGVs that are used to complete tasks such as, the delivery of food, linen, trash, sterile surgical material (instruments and supplies), etc. Each task is assigned a priority level, which changes during the day. This assignment is an effort to balance the use of AGVs. For example, the movement of sterile surgical material from MD to ORD has the highest priority during 3-6pm. Thus, a total of 10 AGVs are dedicated to the delivery of sterile surgical material during this time period.

Every day, MD receives a list of instruments and soft goods that should be delivered to ORD between 3 and 6pm. This list is generated based on doctors' preferences and will be used in surgeries scheduled the next day. Instruments and soft goods are loaded manually into clean case carts. A team of 5 employees is tasked with loading the AGVs. This team is assigned to other tasks during the second shift. Carts are then manually moved to one of the 4 detents available at MD. Detents are areas equipped with the rails necessary for loading and unloading an AGVs. Once a request for an AGV is submitted to AGV control system, an available AGV, closest to the MD, is assigned to the case cart. The case cart is loaded on an AGV. This movement of the AGV is depicted in Figure 1 as "Path of AGV with Clean Cart." To move the case cart to the 2nd floor, this AGV uses elevator J. The clean case cart is then dropped off at one of the 2 detents in the case cart storage area (CCSA). Since CCSA is located next to elevator J on the 2nd floor. The CCSA is not shown in Figure **??**. Since, delivery of food and linen take priority after 6pm, AGVs become increasingly unavailable for the movement of surgical case during those times. Thus, it is expected that the delivery of surgical case carts is completed before other services take priority. The case carts are stored at the CCSA overnight. The case cart is then moved manually to the OR. After the surgery, the soiled cart is moved manually to the detents on the second floor. Once a request for an AGV is submitted to AGV control system, the assigned AGV moves the soiled cart to the CSSD using the path "Path of AGV with Clean and Soiled Cart" depicted in Figure **??**. The soiled instruments are washed and sterilized at the CSSD, a process that takes up to 3 or 4 hours. The sterilized instruments are loaded to a clean case cart and moved to MD for storage. The soiled case carts are washed at the cart washer. The washed case cart is moved to MD for the next cycle. The movement of AGVs with washed case carts is depicted in Figure **??** as "Path of AGVs with Washed Cart".

AGV Scheduling and Operations: The scheduling of AGVs is completed in 2 steps: First, a fleet of AGVs is assigned to tasks during the day based on task priority. Tasks are the delivery of food, delivery of trash, delivery of linen, delivery of surgical carts, etc. Task priorities change during the day. Next, tasks are assigned to AGVs based on a version of the first-come-first-serve rule. For example, if case cart 1 Is ready for pick-up, the 1st available AGV which is located closest to the case cart, will be assigned to deliver the cart.

The operation of AGVs follows certain guidelines, such as, (*i*) AGVs are not allowed to pass each other; (*ii*) if an AGV stops, then other AGVs following will also stop and maintain a safe distance; (*iii*) at most 2 AGVs can use an elevator at the same time; (*iv*) an AGV will not seize elevator J if every detent in the second floor is busy. These operational practices lead to congestion; (*v*) if no task are available, the AGV is moved to the parking area.

Appendix B Appendix Chapter 2

(M1)

$$Z_{\overline{S}} = \min\ c^u\bar{U} + c^o\bar{O} + \sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{it}\bar{X}_{it} \tag{1a}$$

$$t * \bar{Z}_{it} \leq s + \bar{O} \qquad \forall t \in \mathbb{T}, i \in \mathbb{N} \tag{1b}$$

$$s - \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{Z}_{it} \leq \bar{U} \qquad , \tag{1c}$$

$$\sum_{t'=t}^{t+l_i-1}\bar{Z}_{it'} \geq l_i * \bar{X}_{it} \qquad \forall i \in \mathbb{N}, 1 \leq t \leq T - l_i + 1, \tag{1d}$$

$$\sum_{t=1}^{T}\bar{Z}_{it} = l_i \qquad \forall i \in \mathbb{N}, \tag{1e}$$

$$\sum_{i=1}^{N_{dk}}\bar{Z}_{it} \leq 1 \qquad \forall t \in \mathbb{T}, \tag{1f}$$

$$\sum_{t=1}^{T}\bar{X}_{it} = 1 \qquad \forall i \in \mathbb{N}_{dk}, \tag{1g}$$

$$\sum_{t=1}^{T}\bar{X}_{it} \leq 1 \qquad \forall i \in \mathbb{N}, \tag{1h}$$

$$\sum_{t=T-l_i+1}^{T}\bar{X}_{it} = 0 \qquad \forall i \in \mathbb{N}, \tag{1i}$$

$$\bar{Z}_{it} \in \{0,1\}, \qquad \bar{O} \geq 0, \qquad \bar{U} \geq 0, \qquad \bar{X}_{it} \in \{0,1\} \tag{1j}$$

(M2)

$$Z_{\underline{S}} = \min\ c^u\bar{U} + c^o\bar{O} + \sum_{i=1}^{N_{dk}}\sum_{t=1}^{T}\lambda_{it}\bar{X}_{it} \tag{2a}$$

$$t * \bar{Z}_{it} \leq s + \bar{O} \qquad \forall t \in \mathbb{T}, i \in \mathbb{N}_{dk} \tag{2b}$$

$$s - \sum_{i=1}^{N_{dk}}\sum_{t=1}^{T}\bar{Z}_{it} \leq \bar{U} \qquad , \tag{2c}$$

$$\sum_{t'=t}^{t+l_i-1}\bar{Z}_{it'} \geq l_i * \bar{X}_{it} \qquad \forall i \in \mathbb{N}_{dk}, 1 \leq t \leq T - l_i + 1, \tag{2d}$$

$$\sum_{t=1}^{T}\bar{Z}_{it} = l_i \qquad \forall i \in \mathbb{N}_{dk}, \tag{2e}$$

$$\sum_{i=1}^{N_{dk}}\bar{Z}_{it} \leq 1 \qquad \forall t \in \mathbb{T}, \tag{2f}$$

$$\sum_{t=1}^{T}\bar{X}_{it} \leq 1 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2g}$$

$$\sum_{t=T-l_i+1}^{T}\bar{X}_{it} = 0 \qquad \forall i \in \mathbb{N}_{dk}, \tag{2h}$$

$$\bar{Z}_{it} \in \{0,1\}, \qquad \bar{O} \geq 0, \qquad \bar{U} \geq 0, \qquad \bar{X}_{it} \in \{0,1\} \tag{2i}$$

X^*_{idk} represents the assignment of the surgeries to (Day,OR) tuples in (MP). These assignments are denothed by set $\mathbb{N}_{dk}$.

Proposition 4. *The inequality* (2.14a) *is a valid Bender's optimality cut and does not remove any globally integer feasible solution.*

Proof.

- Case 1:

$$\sum_{i\notin\mathbb{N}_{dk}} X_{idk} + \sum_{i\in\mathbb{N}_{dk}}(1 - X_{idk}) = 0$$

$$(Z^{SP}_{dk} - L)(\sum_{i\notin\mathbb{N}_{dk}} X_{idk} + \sum_{i\in\mathbb{N}_{dk}}(1 - X_{idk})) = 0$$

$$\theta_{dk} \geq Z^{SP}_{dk}$$

Thus, $\wedge (X = X^*) \implies \theta_{dk} \geq Z^{SP}_{dk}$.

- Case 2:

$$\sum_{i \notin \mathbb{N}_{dk}} X_{idk} + \sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk}) \geq 1$$

$$(Z_{dk}^{SP} - L)(\sum_{i \notin \mathbb{N}_{dk}} X_{idk} + \sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk})) \leq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)$$

$$= L$$

$$\leq \theta_{dk}$$

Thus, $\wedge (X \neq X^*) \implies \theta_{dk} \geq L$.

□

Here, first case corresponds to all the assignments for a (Day,OR) tuple being the same. In this case, the equation 2.14a reduces to assumed inequality. The second case corresponds to at least one of the assignments for the (Day,OR) tuple under consideration. In this case, inequality 2.14a is redundant by the choice of value of L. The proof of validity for2.16a and 2.15a inequalities is on the same lines.

Proposition 5. *Inequality* (2.15a) *is a valid Bender's optimality cut.*

Proof.

- Case 1:

$$\sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk}) = 0$$

$$(Z_{dk}^{SP} - L)(\sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk})) = 0$$

$$\theta_{dk} \geq Z_{dk}^{SP}$$

Thus, $\wedge (X > X^*) \implies \theta_{dk} \geq Z_{dk}^{SP}$.

- Case 2:

$$\sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk}) \geq 1$$

$$(Z_{dk}^{SP} - L)(\sum_{i \in \mathbb{N}_{dk}} (1 - X_{idk})) \leq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)$$

$$= L$$

$$\leq \theta_{dk}$$

Thus, $\wedge (X \not\geq X^*) \implies \theta_{dk} \geq L$.

□

Here, the first case corresponds to the support of X being a superset of the support of X^* and 2.15a reduces to the assumed inequality. The second case corresponds to the support of X not being a superset of the support of X^*, and 2.15a is redundant by the choice of value of L. The important condition for the inequality 2.15a to be valid is that every superset of the support of X^* must yield $\theta_{dk} \geq Z_{dk}^{SP}$. In other words, adding any surgery to the currently scheduled set of surgeries in a (Day,OR) tuple should not reduce the cost of scheduling for the same.

Proposition 6. *Inequality* (2.16a) *is a valid Bender's optimality cut.*

Proof.

- Case 1:

$$\sum_{i \notin \mathbb{N}_{dk}} X_{idk} = 0$$

$$(Z_{dk}^{SP} - L)(\sum_{i \notin \mathbb{N}_{dk}} X_{idk}) = 0$$

$$\theta_{dk} \geq Z_{dk}^{SP}$$

Thus, $\wedge (X < X^*) \implies \theta_{dk} \geq Z_{dk}^{SP}$.

- Case 2:

$$\sum_{i \notin \mathbb{N}_{dk}} X_{idk} \geq 1$$

$$(Z_{dk}^{SP} - L)(\sum_{i \notin \mathbb{N}_{dk}} X_{idk}) \leq Z_{dk}^{SP} - (Z_{dk}^{SP} - L)$$

$$= L$$

$$\leq \theta_{dk}$$

Thus, $\wedge (X \not< X^*) \implies \theta_{dk} \geq L$.

□

Here, the first case corresponds to the support of X being a subset of the support of X^*, and 2.16a reduces to the assumed inequality. The second case corresponds to the support of X not being a subset of the support of X^*, and 2.16a is redundant by the choice of value of L. The important condition for the inequality 2.16a to be valid is that every subset of the support of X^* must yield $\theta_{dk} \geq Z_{dk}^{SP}$. In other words, removing any surgery from the currently scheduled set of surgeries in a (Day,OR) tuple should not reduce the cost of scheduling for the same.

Bibliography

[1] Cardinal health fourth annual survey finds supply chain tasks causing stress for clinicians and impacting patient care. `https://ir.cardinalhealth.com/news/press-release-details/2019`, May 2019.

[2] S Adler, M Scherrer, KD Rückauer, and FD Daschner. Comparison of economic and environmental impacts between disposable and reusable instruments used for laparoscopic cholecystectomy. *Surgical Endoscopy and Other Interventional Techniques*, 19(2):268–272, 2005.

[3] Ehsan Ahmadi, Dale T Masel, Ashley Y Metcalf, and Kristin Schuller. Inventory management of surgical supplies and sterile instruments in hospitals: a literature review. *Health Systems*, pages 1–18, 2018.

[4] Ehsan Ahmadi, Dale T Masel, Ashley Y Metcalf, and Kristin Schuller. Inventory management of surgical supplies and sterile instruments in hospitals: a literature review. *Health Systems*, 8(2):134–151, 2019.

[5] Ehsan Ahmadi, Dale T Masel, Diana Schwerha, and Seth Hostetler. A bi-objective optimization approach for configuring surgical trays with ergonomic risk consideration. *IISE Transactions on Healthcare Systems Engineering*, 0(just-accepted):1–23, 2019.

[6] Hasan Al Hasan, Christelle Guéret, David Lemoine, and David Rivreau. Surgical case scheduling with sterilising activity constraints. *International Journal of Production Research*, 57(10):2984–3002, 2019.

[7] HUSSEIN HASAN ALI, HENDRIK LAMSALI, and SITI NOREZAM OTHMAN. Operating rooms scheduling using tabu search for elective surgeries in a hospital affected by war-related incidents.

[8] Satyajith Amaran, Nikolaos V Sahinidis, Bikram Sharda, and Scott J Bury. Simulation optimization: a review of algorithms and applications. *Annals of Operations Research*, 240(1):351–380, 2016.

[9] R. Arifin and P. J. Egbelu. Determination of vehicle requirements in automated guided vehicle systems: A statistical approach. *Production Planning and Control*, 11(3):258–270, 2000.

[10] Robert Arifin and Pius J Egbelu. Determination of vehicle requirements in automated guided vehicle systems: a statistical approach. *Production Planning & Control*, 11(3):258–270, 2000.

[11] Roberto Aringhieri, Paolo Landa, Patrick Soriano, Elena Tanfani, and Angela Testi. A two level metaheuristic for the operating room scheduling and assignment problem. *Computers & Operations Research*, 54:21–34, 2015.

[12] AF Attaallah, OM Elzamzamy, AL Phelps, P Ranganthan, and MC Vallejo. Increasing operating room efficiency through electronic medical record analysis. *Journal of perioperative practice*, 26(5):106–113, 2016.

[13] Vincent Augusto, Xiaolan Xie, and Viviana Perdomo. Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers & Industrial Engineering*, 58(2):231–238, 2010.

[14] Jeffrey R Avansino, Adam B Goldin, Renelle Risley, John HT Waldhausen, and Robert S Sawin. Standardization of operative equipment reduces cost. *Journal of pediatric surgery*, 48(9):1843–1849, 2013.

[15] Maya Bam, Brian T Denton, Mark P Van Oyen, and Mark E Cowen. Surgery scheduling with recovery resources. *IISE Transactions*, 49(10):942–955, 2017.

[16] Sakine Batun, Brian T Denton, Todd R Huschka, and Andrew J Schaefer. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS journal on Computing*, 23(2):220–237, 2011.

[17] Smaïl Benzidia, Blandine Ageron, Omar Bentahar, and Julien Husson. Investigating automation and agv in healthcare logistics: a case study based approach. *International Journal of Logistics Research and Applications*, 22(3):273–293, 2019.

[18] Amogh Bhosekar, Sandra Eksioglu, and Tugce Isik. Optimization of automated guided vehicle system in health care facility. In *Proceedings of the 2018 IISE annual conference*, pages 636–641, Orlando, Florida, 2018.

[19] Rafael Calegari, Flavio S Fogliatto, Filipe R Lucini, Michel J Anzanello, and Beatriz D Schaan. Surgery scheduling heuristic considering or downstream and upstream facilities and resources. *BMC Health Services Research*, 20(1):1–11, 2020.

[20] Bruno A Calfa, Anshul Agarwal, Ignacio E Grossmann, and John M Wassick. Hybrid bilevel-lagrangean decomposition scheme for the integration of planning and scheduling of a network of batch plants. *Industrial & Engineering Chemistry Research*, 52(5):2152–2167, 2013.

[21] Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, 119(2):354–366, 2009.

[22] Censis. The challenges of using loaner instruments. `https://censis.com/blog/challenges-using-loaner-instruments/`.

[23] V Chasseigne, G Leguelinel-Blache, TL Nguyen, R de Tayrac, M Prudhomme, JM Kinowski, and P Costa. Assessing the costs of disposable and reusable supplies wasted during surgeries. *International Journal of Surgery*, 53:18–23, 2018.

[24] Chun-Lung Chen and Chuen-Lung Chen. Hybrid metaheuristics for unrelated parallel machine scheduling with sequence-dependent setup times. *The International Journal of Advanced Manufacturing Technology*, 43(1):161–169, 2009.

[25] Mittal Chikul, Hsann Yin Maw, and Yang Kok Soong. Technology in healthcare: A case study of healthcare supply chain management models in a general hospital in singapore. *Journal of Hospital Administration*, 6(6):63–70, 2017.

[26] Christopher P Childers and Melinda Maggard-Gibbons. Understanding costs of care in the operating room. *JAMA surgery*, 153(4):e176233–e176233, 2018.

[27] Nancy Chobin. Best practices for loaner surgical instruments. Sterile Processing University, 2015. http://www.spdceus.com/ceus/pdf/loaner_instrumentation.pdf.

[28] F. F. Choobineh, A. Asef-Vaziri, and X. Huang. Fleet sizing of automated guided vehicles: a linear programming approach based on closed queuing networks. *International Journal of Production Research*, 50(12):3222–3235, 2012.

[29] F Fred Choobineh, Ardavan Asef-Vaziri, and Xiaolei Huang. Fleet sizing of automated guided vehicles: a linear programming approach based on closed queuing networks. *International Journal of Production Research*, 50(12):3222–3235, 2012.

[30] Elvin Coban. The effect of multiple operating room scheduling on the sterilization schedule of reusable medical devices. *Computers & Industrial Engineering*, 147:106618, 2020.

[31] Teodor Gabriel Crainic, Antonio Frangioni, and Bernard Gendron. Bundle-based relaxation methods for multicommodity capacitated fixed charge network design. *Discrete Applied Mathematics*, 112(1-3):73–99, 2001.

[32] Brian Dawson. Five ways to reduce surgical delays. `https://www.whitmanpartners.com/2016/10/01/five-ways-reduce-surgical-delays/`, October 2021.

[33] Renato de Matta. A lagrangean decomposition solution to a single line multiproduct scheduling problem. *European Journal of Operational Research*, 79(1):25–37, 1994.

[34] L. De Vries and R. Huijsman. Supply chain management in health services: an overview. *Supply Chain Management: An International Journal*, 16(3):159–165, 2011.

[35] Latifa Dekhici and Belkadi Khaled. Operating theatre scheduling under constraints. *Journal of Applied Sciences,*, 14(10):1380–1388, 2010.

[36] Mauro Dell'Amico and Marco Trubian. Applying tabu search to the job-shop scheduling problem. *Annals of Operations research*, 41(3):231–252, 1993.

[37] L Demoulin, Katrien Kesteloot, and Freddy Penninckx. A cost comparison of disposable vs reusable instruments in laparoscopic cholecystectomy. *Surgical endoscopy*, 10(5):520–525, 1996.

[38] Brian Denton, James Viapiano, and Andrea Vogl. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science*, 10(1):13–24, 2007.

[39] Brian T Denton, Andrew J Miller, Hari J Balasubramanian, and Todd R Huschka. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research*, 58(4-part-1):802–816, 2010.

[40] G. Dobson, A. Seidman, V. Tilson, and A. Froix. Configuring surgical instrument trays to reduce costs. *IIE Transactions on Healthcare Systems Engineering*, 5(4):225–237, 2015.

[41] Gregory Dobson, Abraham Seidmann, Vera Tilson, and Anthony Froix. Configuring surgical instrument trays to reduce costs. *IIE Transactions on Healthcare Systems Engineering*, 5(4):225–237, 2015.

[42] Twan Dollevoet, J Theresia van Essen, and Kristiaan M Glorie. Solution methods for the tray optimization problem. *European Journal of Operational Research*, 271(3):1070–1084, 2018.

[43] Adam R Dyas, Kelly M Lovell, Courtney J Balentine, Thomas N Wang, John R Porterfield Jr, Herbert Chen, and Brenessa M Lindeman. Reducing cost and improving operating room efficiency: examination of surgical instrument processing. *Journal of Surgical Research*, 229:15–19, 2018.

[44] G Eddie and S White. A comparison of reusable versus disposable laparoscopic instrument costs. *Australian and New Zealand journal of surgery*, 66(10):671–675, 1996.

[45] P. J. Egbelu. The use of non-simulation approaches in estimating vehicle requirements in an automated guided vehicle based transport system. *Material Flow*, 4:17–32, 1987.

[46] Pius J Egbelu. The use of non-simulation approaches in estimating vehicle requirements in an automated guided vehicle based transport system. *Material flow*, 4(1):17–32, 1987.

[47] Burak Ekşioğlu, Sandra Duni Ekşioğlu, and Pramod Jain. A tabu search algorithm for the flowshop scheduling problem with changing neighborhoods. *Computers & Industrial Engineering*, 54(1):1–11, 2008.

[48] Sandra Duni Ekşioğlu, Burak Ekşioğlu, and H Edwin Romeijn. A lagrangean heuristic for integrated production and transportation planning problems in a dynamic, multi-item, two-layer supply chain. *Iie Transactions*, 39(2):191–201, 2007.

[49] Luisa Equi, Giorgio Gallo, Silvia Marziale, and Andres Weintraub. A combined transportation and scheduling problem. *European journal of operational research*, 97(1):94–104, 1997.

[50] Joonyup Eun, Sang-Phil Kim, Yuehwern Yih, and Vikram Tiwari. Scheduling elective surgery patients considering time-dependent health urgency: Modeling and solution approaches. *Omega*, 86:137–153, 2019.

[51] James S Farrelly, Crystal Clemons, Sherri Witkins, Walter Hall, Emily R Christison-Lagay, Doruk E Ozgediz, Robert A Cowles, David H Stitelman, and Michael G Caty. Surgical tray optimization as a simple means to decrease perioperative costs. *Journal of Surgical Research*, 220:320–326, 2017.

[52] Hongying Fei, Chengbin Chu, and Nadine Meskens. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research*, 166(1):91–108, 2009.

[53] Hongying Fei, Nadine Meskens, and Chengbin Chu. An operating theatre planning and scheduling problem in the case of a" block scheduling" strategy. In *2006 International Conference on Service Systems and Service Management*, volume 1, pages 422–428. IEEE, 2006.

[54] Hongying Fei, Nadine Meskens, and Chengbin Chu. A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, 58(2):221–230, 2010.

[55] K. R. Fitzgerald. How to estimate the number of agvs you need. *Modern Materials Handling*, 10:79, 1985.

[56] GI Fragapane, Cevin Zhang, Fabio Sgarbossa, and Jan Ola Strandhagen. An agent-based simulation approach to model hospital logistics. *Int J Simul Model*, 18(4):654–665, 2019.

[57] Giuseppe Ismael Fragapane, Aili Biriita Bertnum, Hans-Henrik Hvolby, and Jan Ola Strandhagen. Material distribution and transportation in a norwegian hospital: A case study. *Ifac-papersonline*, 51(11):352–357, 2018.

[58] Michael C Fu. Optimization via simulation: A review. *Annals of Operations Research*, 53(1):199–247, 1994.

[59] Tharma Ganesharajah, Nicholas G Hall, and Chelliah Sriskandarajah. Design and operational issues in agv-served manufacturing systems. *Annals of Operations Research*, 76:109–154, 1998.

[60] E.. Gebenni, S. Dallari, A. Grassi, G. Perrica, C. Fantuzzi, and R. Gamberini. A simulation based approach for supporting automated guided vehicles (agvs) systems design. *Winter Simulation Conference*, pages 2156–2163, 2008.

[61] Michel Gendreau and Jean-Yves Potvin. Metaheuristics in combinatorial optimization. *Annals of Operations Research*, 140(1):189–213, 2005.

[62] Matthew Gitelis, Yalini Vigneswaran, Michael B Ujiki, Woody Denham, Mark Talamonti, Joseph P Muldoon, and John G Linn. Educating surgeons on intraoperative disposable supply costs during laparoscopic cholecystectomy: a regional health system's experience. *The American Journal of Surgery*, 209(3):488–492, 2015.

[63] Fred Glover. Future paths for integer programming and links to artificial intelligence. *Computers & operations research*, 13(5):533–549, 1986.

[64] Fred Glover. Tabu search: A tutorial. *Interfaces*, 20(4):74–94, 1990.

[65] Fred Glover and Claude McMillan. The general employee scheduling problem. an integration of ms and ai. *Computers & operations research*, 13(5):563–573, 1986.

[66] Fred Glover and Eric Taillard. A user's guide to tabu search. *Annals of operations research*, 41(1):1–28, 1993.

[67] Harry Groenevelt. The just-in-time system. *Handbooks in operations research and management science*, 4:629–670, 1993.

[68] Rosita Guido and Domenico Conforti. A hybrid genetic approach for solving an integrated multi-objective operating room planning and scheduling problem. *Computers & Operations Research*, 87:270–282, 2017.

[69] Monique Guignard. Solving makespan minimization problems with lagrangean decomposition. *Discrete Applied Mathematics*, 42(1):17–29, 1993.

[70] Monique Guignard and Siwhan Kim. Lagrangean decomposition for integer programming: theory and applications. *RAIRO-Operations Research*, 21(4):307–323, 1987.

[71] Alain Guinet and Sondes Chaabane. Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81, 2003.

[72] Diwakar Gupta and Brian Denton. Appointment scheduling in health care: Challenges and opportunities. *IIE transactions*, 40(9):800–819, 2008.

[73] Nicholas G Hall, Chelliah Sriskandarajah, and Tharmarajah Ganesharajah. Operational decisions in agv-served flowshop loops: scheduling. *Annals of Operations Research*, 107(1-4):161–188, 2001.

[74] Lara FB Harvey, Katherine A Smith, and Howard Curlin. Physician engagement in improving operative supply chain efficiency through review of surgeon preference cards. *Journal of minimally invasive gynecology*, 24(7):1116–1120, 2017.

[75] Seyed Hossein Hashemi Doulabi, Louis-Martin Rousseau, and Gilles Pesant. A constraint-programming-based branch-and-price-and-cut approach for operating room planning and scheduling. *INFORMS Journal on Computing*, 28(3):432–448, 2016.

[76] Öncü Hazır and Safia Kedad-Sidhoum. Batch sizing and just-in-time scheduling with common due date. *Annals of Operations Research*, 213(1):187–202, 2014.

[77] Magdy Helal, Ghaith Rabadi, and Ameer Al-Salem. A tabu search algorithm to minimize the makespan for the unrelated parallel machines scheduling problem with setup times. *International Journal of Operations Research*, 3(3), 2006.

[78] Michael Held, Philip Wolfe, and Harlan P Crowder. Validation of subgradient optimization. *Mathematical programming*, 6(1):62–88, 1974.

[79] John N Hooker and Greger Ottosson. Logic-based benders decomposition. *Mathematical Programming*, 96(1):33–60, 2003.

[80] E Huynh, S Klouche, C Martinet, F Le Mercier, T Bauer, and A Lecoeur. Can the number of surgery delays and postponements due to unavailable instrumentation be reduced? evaluating the benefits of enhanced collaboration between the sterilization and orthopedic surgery units. *Orthopaedics & Traumatology: Surgery & Research*, 105(3):563–568, 2019.

[81] Christopher R Jackson, Roland D Eavey, and David O Francis. Surgeon awareness of operating room supply costs. *Annals of Otology, Rhinology & Laryngology*, 125(5):369–377, 2016.

[82] H. Jalali and I. van Nieuwenhuyse. Simulation optimization in inventory replenishment: a classification. *IISE Transactions on Healthcare Systems Engineering*, 47:1217–1235, 2015.

[83] M. Ji and J. Xia. Analysis of vehicle requirements in a general automated guided vehicle system based transportation system. *Computers and Industrial Engineering*, 59(4):544–551, 2010.

[84] RG Kasilingam and SL Gobal. Vehicle requirements model for automated guided vehicle systems. *The International Journal of Advanced Manufacturing Technology*, 12(4):276–279, 1996.

[85] Nikos Katevas. *Mobile robotics in healthcare*, volume 7. IOS Press, 2001.

[86] Sean P Keehan, Gigi A Cuckler, John A Poisal, Andrea M Sisko, Sheila D Smith, Andrew J Madison, Kathryn E Rennie, Jacqueline A Fiore, and James C Hardesty. National health expenditure projections, 2019–28: Expected rebound in prices drives rising spending growth: National health expenditure projections for the period 2019–2028. *Health Affairs*, 39(4):704–714, 2020.

[87] Jeff Kennington. The fixed-charge transportation problem: a computational study with a branch-and-bound code. *AIIE Transactions*, 8(2):241–247, 1976.

[88] Marwa Khalfalli, Fouad Ben Abdelaziz, and Hichem Kamoun. Multi-objective surgery scheduling integrating surgeon constraints. *Management Decision*, 2019.

[89] Wieslaw Kubiak, George Steiner, and Julian Scott Yeomans. Optimal level schedules for mixed-model, multi-level just-in-time assembly systems. *Annals of Operations Research*, 69:241–259, 1997.

[90] Gobal S. L. and R. G. Kasilingam. A simulation model for estimating vehicle requirements in automated guided vehicle systems. *Computers and Industrial Engineering*, 21(1):623–627, 1991.

[91] Guillermo Latorre-Núñez, Armin Lüer-Villagra, Vladimir Marianov, Carlos Obreque, Francisco Ramis, and Liliana Neriz. Scheduling operating rooms with consideration of all resources, post anesthesia beds and emergency surgeries. *Computers & Industrial Engineering*, 97:248–257, 2016.

[92] Tuan Le-Anh and MBM De Koster. A review of design and control of automated guided vehicle systems. *European Journal of Operational Research*, 171(1):1–23, 2006.

[93] J. Lee, R. H.-G. Choi, and M. Khaksar. Evaluation of automated guided vehicle systems by simulation. *Computers and Industrial Engineering*, 19(1):318–321, 1910.

[94] Sangbok Lee and Yuehwern Yih. Reducing patient-flow delays in surgical suites through determining start-times of surgical cases. *European Journal of Operational Research*, 238(2):620–629, 2014.

[95] Xiangyong Li, Lanjian Zhu, Fazle Baki, and AB Chaouch. Tabu search and iterated local search for the cyclic bottleneck assignment problem. *Computers & Operations Research*, 96:120–130, 2018.

[96] Rasaratnam Logendran, Brent McDonell, and Byran Smucker. Scheduling unrelated parallel machines with sequence-dependent setups. *Computers & Operations Research*, 34(11):3420–3438, 2007.

[97] B. Mahadevan and T. T. Narendran. Estimation of number of agvs for an fms: an analytical model. *International Journal of Production Research*, 31(7):1655–1670, 1993.

[98] Dimitrios K Manatakis and Nikolaos Georgopoulos. Reducing the cost of laparoscopy: reusable versus disposable laparoscopic instruments. *Minimally invasive surgery*, 2014, 2014.

[99] Suneil Mandava. How much is consignment costing your hospital?, 2017. http://www.mobileaspects.com/how-much-is-consignment-costing-your-hospital.

[100] R. J. Mantel and H. R. A. Landeweerd. Design and operational control of an agv system. *International Journal of Production Economics*, 41(1):257–266, 1995.

[101] Eric Marcon and Franklin Dexter. Impact of surgical sequencing on post anesthesia care unit staffing. *Health care management science*, 9(1):87–98, 2006.

[102] Inês Marques, M Eugénia Captivo, and Margarida Vaz Pato. An integer programming approach to elective surgery scheduling. *OR spectrum*, 34(2):407–427, 2012.

[103] W. R. Maxwell and J. A. Muckstadt. Design of automatic guided vehicle systems. *AIIE Transactions*, 14(2):114–124, 1982.

[104] William L Maxwell and Jack A Muckstadt. Design of automatic guided vehicle systems. *IIE Transactions*, 14(2):114–124, 1982.

[105] Garth P McCormick. Nonlinear programming; theory, algorithms, and applications. Technical report, 1983.

[106] B. Meijboom, S. Schmidt-Bakx, and G. Westert. Supply chain management practices for improving patient-oriented care. *Supply Chain Management: An International Journal*, 16(3):166–175, 2011.

[107] Nadine Meskens, David Duvivier, and Arnauld Hanset. Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems*, 55(2):650–659, 2013.

[108] Philippe Michelon, Marcelo Dib Cruz, and Viviane Gascon. Using the tabu search method for the distribution of supplies in a hospital. *Annals of Operations Research*, 50(1):427–435, 1994.

[109] Karen Moons, Geert Waeyenbergh, and Liliane Pintelon. Measuring the logistics performance of internal hospital supply chains–a literature study. *Omega*, 82:205–217, 2019.

[110] Amirhossein Moosavi and Onur Ozturk. Metaheuristics for the operating theater planning and scheduling: A systematic review. *arXiv preprint arXiv:2008.04970*, 2020.

[111] Dante Mora-Mariano, Miguel Angel Gutiérrez-Limón, and Antonio Flores-Tlacuahuac. A lagrangean decomposition optimization approach for long-term planning, scheduling and control. *Computers & Chemical Engineering*, 135:106713, 2020.

[112] Eric Muñoz, William Muñoz III, and Leslie Wise. National and surgical health care expenditures, 2005–2025. *Annals of surgery*, 251(2):195–200, 2010.

[113] Kelly Nast and KA Swords. Decreasing operating room costs via reduction of surgical instruments. *Journal of pediatric urology*, 15(2):153–e1, 2019.

[114] NHEA. *National Health Expenditure Accounts:Methodology Paper, 2017*, 2017. `https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/dsm-17.pdf` [Accessed: Whenever].

[115] L. Niziolek, T. C. Chiam, and Y. Yih. A simulation-based study of distribution strategies for pharmaceutical supply chains. *IISE Transactions on Healthcare Systems Engineetring*, 2:181–189, 2012.

[116] Ibrahim H Osman and Gilbert Laporte. Metaheuristics: A bibliography, 1996.

[117] Selin Özpeynirci, Burak Gökgür, and Brahim Hnich. Parallel machine scheduling with tool loading. *Applied Mathematical Modelling*, 40(9-10):5660–5671, 2016.

[118] Udatta S Palekar, Mark H Karwan, and Stanley Zionts. A branch-and-bound method for the fixed charge transportation problem. *Management science*, 36(9):1092–1105, 1990.

[119] M. Pedan, M. Gregor, and D. Plinta. Implementation of automated guided vehicle system in healthcare facility. *Procedia Engineernting*, 192(Supplement C):665–670, 2017.

[120] Marko Pedan, Milan Gregor, and Dariusz Plinta. Implementation of automated guided vehicle system in healthcare facility. *Procedia engineering*, 192:665–670, 2017.

[121] Marion L Penn, Chris N Potts, and Paul R Harper. Multiple criteria mixed-integer programming for incorporating multiple factors into the development of master operating theatre timetables. *European Journal of Operational Research*, 262(1):194–206, 2017.

[122] Juta Pichitlamken and Barry L Nelson. A combined procedure for optimization via simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 13(2):155–179, 2003.

[123] José Luís Quesado Pinto, João Carlos O Matias, Carina Pimentel, Susana Garrido Azevedo, and Kannan Govindan. Introduction to lean and just-in-time manufacturing. In *Just in Time Factory*, pages 1–4. Springer, 2018.

[124] Sergey Polyakovskiy, Dhananjay Thiruvady, and Rym M'Hallah. Just-in-time batch scheduling subject to batch size. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 228–235, 2020.

[125] S. Rajotia, K. Shanker, and J. L. Batra. Determination of optimal agv fleet size for an fms. *International Journal of Production Research*, 36(5):1177–1198, 1998.

[126] S Rajotia, K Shanker, and JL Batra. Determination of optimal agv fleet size for an fms. *International Journal of Production Research*, 36(5):1177–1198, 1998.

[127] Francis Reymondon, Bertrand Pellet, and Eric Marcon. Optimization of hospital sterilization costs proposing new grouping choices of medical devices into packages. *International Journal of Production Economics*, 112(1):326–335, 2008.

[128] Stewart Robinson. Conceptual modelling for simulation part ii: a framework for conceptual modelling. *Journal of the operational research society*, 59(3):291–304, 2008.

[129] Manuel D Rossetti, Robin A Felder, and Amit Kumar. Simulation of robotic courier deliveries in hospital distribution services. *Health care management science*, 3(3):201–213, 2000.

[130] Manuel D Rossetti and Francesco Selandari. Multi-objective analysis of hospital delivery systems. *Computers & Industrial Engineering*, 41(3):309–333, 2001.

[131] Alireza Saremi, Payman Jula, Tarek ElMekkawy, and G Gary Wang. Appointment scheduling of outpatient surgical services in a multistage operating room department. *International Journal of Production Economics*, 141(2):646–658, 2013.

[132] Robert G Sargent. Verification and validation of simulation models. In *Proceedings of the 2010 Winter Simulation Conference*, pages 166–183. IEEE, 2010.

[133] Gabriel N Schaer, Ossi R Koechli, and Urs Haller. Single-use versus reusable laparoscopic surgical instruments: a comparative cost analysis. *American journal of obstetrics and gynecology*, 173(6):1812–1815, 1995.

[134] Rose Seavey. Reducing the risks associated with loaner instrumentation and implants. *AORN journal*, 92(3):322–334, 2010.

[135] Hanif D Sherali and J Cole Smith. Improving discrete model representations via symmetry considerations. *Management Science*, 47(10):1396–1407, 2001.

[136] D. Sinriech and J. M. A. Tanchoco. An economic model for determining agv fleet size. *International Journal of Production Research*, 30(6):1255–1267, 1992.

[137] D Sinriech and JMA Tanchoco. An economic model for determining agv fleet size. *International Journal of Production Research*, 30(6):1255–1268, 1992.

[138] David E Skarda, Michael Rollins, Seth Andrews, Molly McFadden, Doug Barnhart, Rebecka Meyers, and Eric Scaife. One hospital, one appendectomy: the cost effectiveness of a standardized doctor's preference card. *Journal of pediatric surgery*, 50(6):919–922, 2015.

[139] Eunhye Song and Barry L Nelson. Input–output uncertainty comparisons for discrete optimization via simulation. *Operations Research*, 67(2):562–576, 2019.

[140] Emily Walker Stockert and Alexander Langerman. Assessing the magnitude and costs of intraoperative inefficiencies attributable to surgical instrument trays. *Journal of the American College of Surgeons*, 219(4):646–655, 2014.

[141] Simon Strassl and Nysret Musliu. Instance space analysis and algorithm selection for the job shop scheduling problem. *Computers & Operations Research*, page 105661, 2022.

[142] J. M. A. Tanchoco, P. J. Egbelu, and F. Taghaboni. Determination of the total number of vehicles in an agv-based material transport system. *Material FLow*, 4:33–51, 1987.

[143] JMA Tanchoco, PJ Egbelu, and Fataneh Taghaboni. Determination of the total number of vehicles in an agv-based material transport system. *Material flow*, 4(1):33–51, 1987.

[144] Tecsys. Hidden costs of logistics in healthcare. `https://infohub.tecsys.com`, May 2019.

[145] Sebastian Terrazas-Moreno and Ignacio E Grossmann. A multiscale decomposition method for the optimal planning and scheduling of multi-site continuous multiproduct plants. *Chemical Engineering Science*, 66(19):4307–4318, 2011.

[146] Angela Testi, Elena Tanfani, and Giancarlo Torre. A three-phase approach for operating theatre schedules. *Health care management science*, 10(2):163–172, 2007.

[147] Joris Van de Klundert, Philippe Muls, and Maarten Schadd. Optimizing sterilization logistics in hospitals. *Health care management science*, 11(1):23–33, 2008.

[148] Mary M Van Meter and Rony A Adam. Costs associated with instrument sterilization in gynecologic surgery. *American journal of obstetrics and gynecology*, 215(5):652–e1, 2016.

[149] Wim Vancroonenburg, Pieter Smet, and Greet Vanden Berghe. A two-phase heuristic approach to multi-day surgical case scheduling considering generalized resource constraints. *Operations Research for Health Care*, 7:27–39, 2015.

[150] François Vanderbeck and Laurence A Wolsey. Reformulation and decomposition of integer programs. In *50 Years of Integer Programming 1958-2008*, pages 431–502. Springer, 2010.

[151] Mohsen Varmazyar, Raha Akhavan-Tabatabaei, Nasser Salmasi, and Mohammad Modarres. Operating room scheduling problem under uncertainty: Application of continuous phase-type distributions. *IISE Transactions*, 52(2):216–235, 2020.

[152] Bharathwaj Vijayakumar, Pratik J Parikh, Rosalyn Scott, April Barnes, and Jennie Gallimore. A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *European Journal of Operational Research*, 224(3):583–591, 2013.

[153] I. F. A. Vis. Survey of research in the design and control of automated guided vehicle systems. *European Journal of Operational Research*, 170(3):677–709, 2006.

[154] Iris FA Vis. Survey of research in the design and control of automated guided vehicle systems. *European Journal of Operational Research*, 170(3):677–709, 2006.

[155] Iris FA Vis, René De Koster, Kees Jan Roodbergen, and Leon WP Peeters. Determination of the number of automated guided vehicles required at a semi-automated container terminal. *Journal of the Operational research Society*, 52(4):409–417, 2001.

[156] Jonas Volland, Andreas Fügener, Jan Schoenfelder, and Jens O Brunner. Material logistics in hospitals: a literature review. *Omega*, 69:82–101, 2017.

[157] Stefan Voß. Meta-heuristics: The state of the art. In *Workshop on Local Search for Planning and Scheduling*, pages 1–23. Springer, 2000.

[158] AJ Weiss, A Elixhauser, and RM Andrews. Characteristics of operating room procedures in us hospitals, 2011: statistical brief# 170. 2006.

[159] Mike Williams. Materials management and logistics in the emergency department. *Emergency Medicine Clinics*, 22(1):195–215, 2004.

[160] Laurence A Wolsey and George L Nemhauser. *Integer and combinatorial optimization*, volume 55. John Wiley & Sons, 1999.

[161] Dan Wu and Marianthi G Ierapetritou. Decomposition approaches for the efficient solution of short-term scheduling problems. *Computers & Chemical Engineering*, 27(8-9):1261–1276, 2003.

[162] I Wubben, Jeanette Gabrielle van Manen, BJ Van den Akker, SR Vaartjes, and Willem H van Harten. Equipment-related incidents in the operating room: an analysis of occurrence, underlying causes and consequences for the clinical process. *Qual Saf Health Care*, 19(6):e64–e64, 2010.

[163] Haokun Yang, David E Bernal, Robert E Franzoi, Faramroze G Engineer, Kysang Kwon, Sechan Lee, and Ignacio E Grossmann. Integration of crude-oil scheduling and refinery planning by lagrangean decomposition. *Computers & Chemical Engineering*, 138:106812, 2020.

[164] X. Yang. Metaheuristic Optimization. *Scholarpedia*, 6(8):11472, 2011. revision #91488.

[165] T. Yifei, C. Junruo, L. Meihong, L. Xianxi, and F. Yali. An estimate and simulation approach to determining the automated guided vehicle fleet size in fms. *3rd International Conference on Computer Science and Information Technology*, 9:432–435, 2010.

[166] Shuwan Zhu, Wenjuan Fan, Shanlin Yang, Jun Pei, and Panos M Pardalos. Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization*, 37(3):757–805, 2019.