Stochastic Models for the Operating Room Scheduling Problem under Uncertainty

by
Amirhossein Najjarbashi

A Dissertation submitted to the Department of Industrial Engineering,

Cullen College of Engineering

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Industrial Engineering

Chair of Committee: Gino J. Lim

Committee Member: Jiming Peng

Committee Member: Taewoo Lee

Committee Member: Edward Kao

Committee Member: Ming Zhao

University of Houston
May 2020

Dedication

To my one and only, Mahtab, whose love empowers me to achieve impossible

To my family, for all their support and love from thousands of miles away

To every underdog, who dreams big and works hard to beat the odds

Acknowledgements

It is a pleasure to thank those who made this dissertation possible. First and foremost, I offer my deepest gratitude to my advisor, Dr. Gino Lim, for his immense knowledge, significant technical insights, constant advice, and encouragement throughout this program. It was a great honor to be his student. Many thanks to the dissertation committee members, Dr. Taewoo Lee, Dr. Jiming Peng, Dr. Ming Zhao and Dr. Edward Kao for providing valuable comments that improved the quality of my work.

My sincere thanks to my better half, companion and best friend, Mahtab, whose unconditional love and support made this possible. Abundant thanks to my wonderful friends, Nima, Elham, Bahar, Ali, Aein, Azin, Aida and Amirali who became my US family and made the past six years an era full of sweet memories. Houston will always be in my heart. I am thankful to my friends and colleagues, Navid, Anahita, Maryam, Mohammad, Zahed and Shiva for their guidance and support, which paved my way to a successful Ph.D.

Abstract

Surgical procedures are complex tasks requiring a variety of specialized and expensive resources. They are recognized amongst the most crucial activities in hospitals from social, medical and economic points of view. Accordingly, operating room (OR) management has gained significant attention within the past decade. Specifically, efficient planning and scheduling of operating rooms has become a priority for both practitioners and researchers. Planning problems deal with assigning surgical patients to certain days over a given planning horizon (e.g., a week) considering limited resources (e.g., PACU bed). Scheduling problems address the allocation of patients to the OR and the calculation of surgery start times on a daily basis. One of the major problems associated with the development of accurate OR planning and scheduling strategies is the uncertainty inherent to surgical procedures. Among many sources of uncertainty mentioned in the literature, high variability of surgery durations is identified as the biggest challenge towards developing practical OR schedules. Uncertain surgery durations can cause a large deviation from the expected completion time of all surgery cases scheduled for each day. Larger than average surgery durations cause an extended overtime for surgical teams and waiting time for patients. On the other hand, shorter than expected durations result in unnecessary resource idle time and lost revenue. The goal of the proposed research is to address surgery duration uncertainty and make a trade-off between optimizing average performance and reducing variability of the performance measures via developing risk-based models and solution methods.

A risk averse solution method using Conditional Value-at-Risk (CVaR) is proposed to reduce variability on overtime and its associated costs in a daily OR scheduling problem. CVaR is a risk measure that is shown to be effective in simultaneously reducing the expected value and the variance of a performance measure. The OR scheduling problem is formulated as a stochastic mixed-integer linear programming model, where a surgery duration follows a known probability distribution function. Numerical results from real-life instances show that our approach not only outperforms the widely used expected value (EV) approach in reducing variability on overtimes, but it also shows acceptable performance in minimizing

the expected value of overtimes and idle times. We also show that the proposed method performs well under distributional uncertainty of the random data. As compared to the EV in terms of the total cost, the CVaR reduced the variance, interquartile range and median absolute deviation by 37%, 25% and 24%, respectively, with a slight increase (4%) in the expected value. The proposed method does not explicitly optimize the average performance but focuses on a rather small portion of scenarios. Our next work tries to tackle this limitation using chance-constrained programming.

This work proposes a two-stage chance-constrained model to solve the OR scheduling problem under uncertainty. The goal is to minimize the costs associated with OR opening and overtimes and patient waiting times. The risk of OR overtime is controlled using individual chance constraints. A deterministic equivalent formulation of the model is proposed. Numerical experiments on several test problems show that the proposed model provides a better trade-off between minimizing costs and reducing solution variability compared to two existing models in the literature. We used several criteria such as OR utilization, amount of OR overtime, patient waiting time and the ratio of overtime scenarios to compare the three models. We exploit the structure of the model in order to propose a decomposition algorithm that solves large test instances of the OR scheduling problem, that of which is known to be NP-hard. Strong valid inequalities are derived in order to accelerate the convergence speed. It is shown that the proposed algorithm will achieve global optimality. Moreover, the proposed algorithm outperforms a commercial solver and a basic decomposition algorithm by solving instances of up to double size to optimality within one hour. Specifically, this algorithm solves instances with up to 74 surgeries and 20 ORs to optimality. The next chapter aims to develop more powerful solution techniques to solve even larger instances of the stochastic OR scheduling problem.

Lower- and upper-bounds are derived for the two-stage model using Lagrangian relaxation and CVaR. An augmented decomposition algorithm is proposed that is able to find high quality initial feasible solutions in reasonable solution times. Numerical experiments show that the proposed bounds enhance the computational efficiency of the decomposition algorithm significantly. The augmented algorithm is able to solve instances with up to

207 surgeries and 40 ORs to optimality within one hour. As the probability threshold for OR overtime increases, the proposed algorithm provides stronger bounds on the optimal objective value.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Health care expenditures are expected to constitute 25% of the US gross domestic product (GDP) in 2025, up from 15.9% in 2005 [66]. Operating rooms contribute to more than 30% of total expenses [56] and 40% of total revenues [33] in hospitals. The rapid growth of healthcare expenditures adds to the significance of proper operating room management. This increase is fueled by new technologies, new medications, aging population and the shortage of skilled staff in hospitals [51].

Surgical expenses contribute to 30% of health care expenditures, and are expected to grow from \$572 billion in 2005 to \$912 billion (2005 dollars) in the year 2025 (7.3% of US GDP). Surgical procedures are complex tasks requiring a variety of specialized and expensive resources. Weiss, Elixhauser and Andrews [94] reported that in 2011, hospitalizations that involved surgical procedures constituted 29% of the total hospital stays and 48% of the total hospital costs in the US. Hospital stays that involved a surgical procedure were about twice as costly as other hospital stays. In light of these reports, surgeries are recognized amongst the most crucial activities in hospitals from social, medical and economic points of view. Therefore, operating room (OR) management has gained significant attention within the past decade. Specifically, efficient planning and scheduling of operating rooms has become a priority for both practitioners and researchers.

1.2 Operating Room Scheduling

An operating room refers to a medical facility designed and equipped to perform surgeries. Cardoen, Demeulemeester and Beliën [12] categorized the operating room scheduling process into four planning levels. The first stage, often regarded as case mix planning (CMP), determines how much operating room time is assigned to the different surgeons or surgical groups in the long run. This stage is situated on a strategic level. The second stage, which is tactically oriented, deals with the development of a master surgery schedule (MSS). This schedule determines the number and type of operating rooms available, the operating time limits for the available rooms, and the surgeons or surgical groups to whom the operating room time is assigned. Our research is focused on the third stage, where surgical cases are scheduled on a daily basis. A two-step approach is commonly applied to address this stage. In the first step, patients are assigned to different days in the planning horizon (e.g., a week). In the second step, the patient population for a specific day is assigned to the available operating rooms and is sequenced. This stage will be explained in further detail later in this chapter. Finally, the execution of the surgery schedule is monitored online in the fourth stage. When uncertainties realize and the surgery schedule is substantially disrupted, rescheduling may be necessary. Figure 1.1 shows the general workflow of a surgical procedure in an operating room. A variety of activities are performed through pre-operative, peri-operative and post-operative stages. The pre-operative stage begins with the surgery decision and continues until the patient's arrival to the OR. It typically includes all preparations prior to the surgery such as physical examinations, medical tests, and administrative work. The peri-operative stage includes all activities performed in the OR and ends with the patient's transfer to the recovery area. The post-operative stage includes recovery and follow-up periods. Our research is focused on the activities performed during the peri-operative stage and the resources that are involved during this stage.

Samudra, Van Riet, Demeulemeester, Cardoen, Vansteenkiste and Rademakers [79] broke down the daily OR scheduling problem using seven descriptive fields: patient characteristics, performance measures, decision delineation, upstream and downstream facilities,

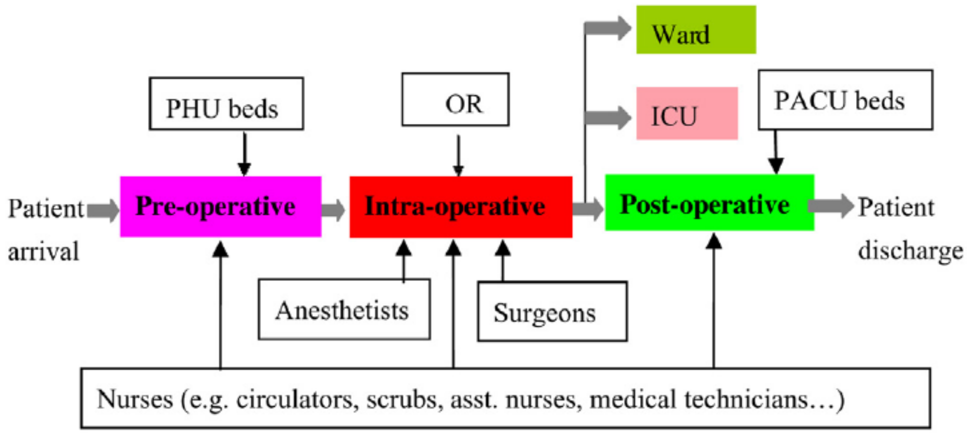


Figure 1.1: General Workflow of a Surgical Case in the OR [95]

uncertainty, operations research methodology, and testing phase and application.

- **Patient types**: Two major patient classes are considered in the literature: elective patients and non-elective patients. Elective patients are those for whom the surgery can be planned in advance (e.g., weeks, months). On the other hand, non-elective patients arrive randomly to the hospital and their surgeries must be fitted into the schedule on short notice. Elective patients are further divided into inpatients and outpatients. Inpatients are hospitalized patients who are admitted to the hospital following a doctor's order and have to stay overnight. Outpatients are those patients who are not officially admitted to the hospital and typically enter and leave the hospital on the same day. There are some features that distinguishes an outpatient from an inpatient. For instance, outpatient surgeries often include more standardized procedures such as routine surgeries and minimally invasive procedures. Moreover, the actual arrival time of outpatients is uncertain because they are not admitted to the hospital in advance. A non-elective surgery is commonly considered an *emergency* if it has to be performed immediately and an *urgency* if it can be postponed for a short time (i.e., days). Non-elective surgeries are often scheduled in two ways: 1) They are incorporated in the elective surgery scheduling by reserving buffer OR time, or 2) A number of ORs are exclusively assigned to the non-elective patients such that these operations take place with minimum waiting time.

- **Performance criteria**: Choosing the best performance measure depends on the angle from which we are looking at the problem and what our priorities are. If we are hospital administrators, achieving high utilization levels and low costs is the most important goal. On the other hand, surgeons and medical staff care less about costs and revenues and would rather have their working preferences met instead (i.e., less overtime and less night shifts). The surgical patients, who are the client of the hospital, desire short waiting times and no postponed surgeries. The most common criteria considered in the literature are waiting time, utilization, leveling, idle time, throughput, preferences, financial measures, makespan, and patient deferral. Many articles in the literature tried to address the interests of different stakeholders by considering a multitude of criteria. They often assign an importance weight to each measure and optimize the weighted sum of each. Waiting time is perceived as both direct (i.e., waiting time on the day of surgery to start the operation) and indirect (i.e., time spent on the waiting list) waiting time. It is common to consider underutilization as undertime and overutilization as overtime, although this might not be the case. Utilization refers to the workload of a resource, while undertime or overtime are timing aspects. Therefore, we can have an underutilized OR which runs into overtime. Minimizing overtime is one of the most favorable performance measures due to its huge negative consequences, such as job dissatisfaction, high costs, and surgery cancellations. Preference-based performance measures are often related to qualitative aspects desired by either the medical staff or the patients. For example, surgeons prefer to have their surgical patients scheduled on their selected date and at a single facility.

- **Decision types**: The decisions considered in the OR planning and scheduling problems can be categorized based on three main groups: 1) discipline level, 2) surgeon level, and 3) patient level. The most common decisions are date, room, start time, and resource capacity (e.g., OR time blocks, PACU beds). In the discipline level, decisions are made for a whole medical department. In the surgeon level, decisions involve a surgeon or a whole surgical group with the same specialty. Similar rule applies to the decisions made for the surgical patients. A large portion of the published literature deals with assigning dates and rooms to individual patients or patient types. A very important distinction when defining the decisions is the OR scheduling policy. In the *block* scheduling policy, OR time slots are dedicated to a discipline or to a surgeon group. Therefore, surgeons are only allowed to book cases into the blocks assigned to them. On the other hand, the *open* scheduling policy relaxes this restriction and allows surgeons to book their surgeries in any available time slots. Many articles in the literature have reported that the majority of hospitals are conventionally using the block booking policy.

- **Upstream and downstream facilities**: We mentioned earlier that a large portion of hospital admissions belong to surgical patients. Therefore, many departments in a hospital are involved in surgical procedures from the pre-operative stage to the post-operative stage. Operating room scheduling affects the upstream and downstream resources such as ward, PACU and ICU. Therefore, devising integrated models while taking into account upstream and downstream resources can improve the overall performance of a hospital. Many articles in the literature have considered the PACU capacity in their models, while the other resources have been largely ignored by the research community.

- **Uncertainty**: One of the major problems associated with the development of accurate OR planning and scheduling strategies is the uncertainty inherent to surgical services. Deterministic planning and scheduling approaches ignore uncertainty, whereas stochastic approaches explicitly incorporate it. Stochasticity in the form of uncertain patient arrivals and surgery durations is frequently incorporated. Non-elective patient arrivals are in most cases impossible to predict in advance and additionally occupy a random amount of OR time, which often leaves OR managers with no option but to reserve capacity for them. In contrast, the arrival of elective patients to ORs contains little uncertainty and is frequently considered as deterministic in the literature. Surgery durations are difficult to predict because the magnitude of the procedure for some surgeries only becomes apparent once the surgery is already in progress. Additionally, the durations often depend on various complex factors, e.g., the characteristics of the patient, the surgeon and the surgical team. As individual surgery durations are uncertain, their sum, or the total workload per OR, is also uncertain (See Figure 1.2). Surgery rescheduling limits the impact that deviations from the initial OR schedule have on the hospital. These deviations on the day of surgery are caused by an uncertain workload due to possible emergency arrivals, deviations from the estimated surgery durations or variable LOS in downstream units. Other causes that can lead to deviations include staff unavailability, equipment failure, late arrival of patients or staff and, in an outpatient setting, patient no-shows. To limit the impact, interventions throughout the day in the form of rescheduling might be needed. Two main types of interventions are cancellations and OR reassignments. In the case of an OR reassignment, the patient is still served on the planned day, but is moved or rescheduled to another OR. A more severe intervention is when a patient cannot be served on the planned day and needs to have the assignment canceled. This patient will need to be fit into the elective schedule of another day.

- **Operations research methodology**: Throughout the decades of studying OR scheduling problems, a wide variety of operations research techniques have been employed. Simulation techniques, mathematical programming models, improvement heuristics and scenario analysis are amongst the most commonly used approaches. A significant increase in the application of mathematical programming models can be observed from the beginning of the 21$^{\text{st}}$ century. Linear programming, goal programming, and integer and mixed integer programming are widely used by the operations research community to model the OR scheduling problems.

- **Testing phase and application**: Researchers have used both theoretical and real-world data to evaluate the applicability of their models. Many articles have used data collected from hospitals to show that their proposed approach can improve some

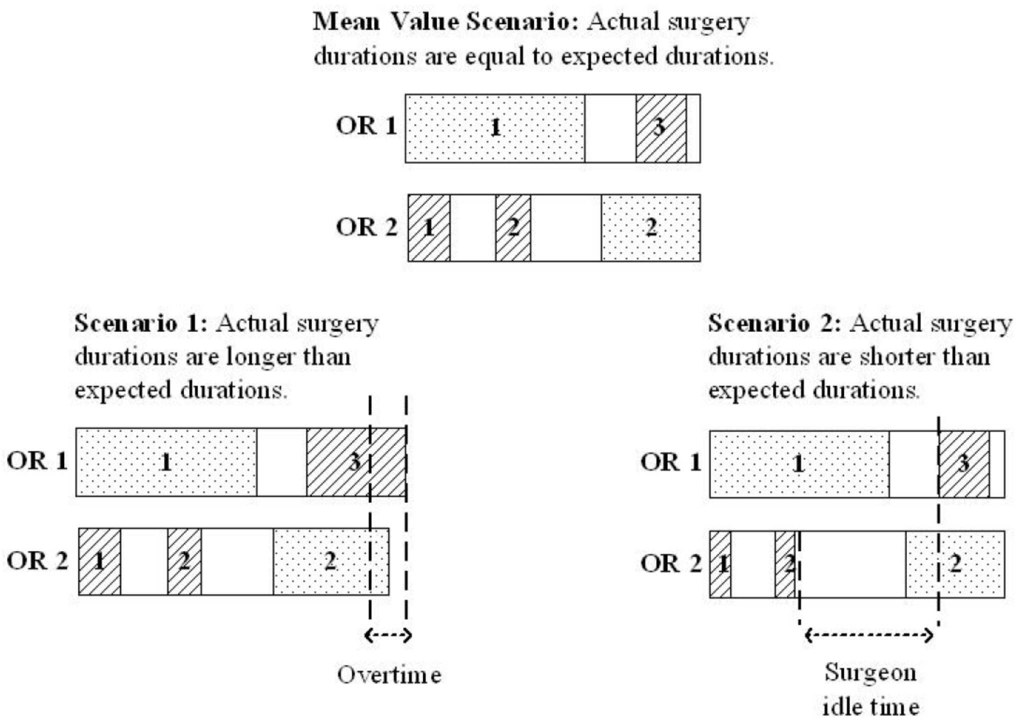


Figure 1.2: The Effect of Surgery Duration Uncertainty on OR Overtime and Surgeon Idle Time: A Simple Example [4]

aspects of performance such as utilization, overtime and waiting times. However, it is important to note that the implementation of these models in practice might lead to completely different outcomes than expected since there are a large number of factors that are neglected when modeling the OR scheduling problems. Moreover, the obtained results from one hospital may not be necessarily applicable to other facilities, as each facility has its own rules and policies.

This research focuses on the second step of the OR scheduling problem, i.e., assigning the patient population for each day to the available ORs and scheduling cases in each OR. Elective patients are the only patient type considered in our study assuming that a number of ORs are dedicated to non-elective patients. The fixed cost of opening ORs, the expected costs associated with OR overtime and idle time, and the patient waiting time are the performance measures to be minimized in the proposed models in this research. The proposed models in our work make decisions regarding OR opening, surgery to OR assignment, and projected start time. Operating rooms are the only resources considered in our study. It is assumed that the upstream and downstream resources are sufficiently available and do not restrict our scheduling problem. Uncertain surgery durations are taken into account in our research. It is assumed that surgery durations follow known probability distributions. The underlying distribution is approximated by a discrete and finite support using the Monte Carlo sampling method. We propose three types of stochastic mixed-integer programming (MIP) models for the OR scheduling problem in our research. The effectiveness of the proposed models and solution methods is tested using several criteria on a variety of real-life and theoretical test problem instances borrowed from the literature.

1.3 Optimization under Uncertainty

Optimization under uncertainty refers to a class of optimization problems where there are uncertainties involved in the data or the model. A key challenge in optimization under uncertainty is when we are dealing with a huge uncertainty space that leads to large-scale optimization models. Solving optimization problems under uncertainty becomes further complicated in the presence of discrete decision variables (e.g., binary, integer) that model logical and combinatorial decisions. This section briefly introduces the theory and methodology that have been applied to solve the OR scheduling problems under uncertainty. We discuss and contrast the widely used approaches in the literature, namely recourse-based stochastic programming and robust optimization. Furthermore, we introduce some of the risk-averse approaches including chance-constrained programming, and discuss how they can benefit operating rooms in the presence of variability in the problem parameters.

Many articles in the literature have used the two-stage stochastic programming approach to model the OR scheduling problem under uncertainty. Under the standard two-stage stochastic programming paradigm, the decision variables are partitioned into two sets. The first-stage variables have to be decided before the actual realization of the uncertain parameters. Subsequently, once we know the value of the random parameters, recourse actions can be made in the second-stage problem. Due to uncertainty, the objective function of the second-stage problem is a random variable. The objective is to choose the first-stage variables in a way that the sum of the first-stage costs and the *expected value* of the random second-stage costs is minimized. The standard formulation for the two-stage stochastic mixed-integer programming model can be shown as [83]

$$\begin{aligned} Min \quad & c^T x + E_{\xi \in \Omega}[Q(x, \xi)], \\ & \text{subject to} \\ & Ax = b, \\ \text{and} \quad & x \in X \end{aligned} \tag{1.3.1}$$

where the set X represents the integrality constraints, and $Q(x, \xi)$ is the optimal value of the second-stage problem.

$$\begin{aligned} \underset{y \in \Re_+^{n2}}{Min} \quad & q(\xi)^T y, \\ & \text{subject to} \\ T(\xi)x + W(\xi)y = h(\xi) \end{aligned} \tag{1.3.2}$$

where $c \in R^{n_2}$ and ξ is a random variable from a probability space (Ω, F, P) with $\Omega \in R^k$, $f : \Omega \to R^{n_2}$, $h : \Omega \to R^{m_2}$, $W : \Omega \to R^{m_2 \times n_2}$, $T : \Omega \to R^{m_2 \times n_1}$. The sample average approximation (SAA) method [44] is a widely used approach to solve the two-stage stochastic programs. To use this approach, it is assumed that the expectation function is well defined and finite-valued for all $y \in \Re_+^{n2}$. This implies that the value of the second-stage cost is finite for every realization of ξ. Suppose that we have a sample $\xi_1, \xi_2, ..., \xi_N$ of N realizations of the random vector ξ. This sample can be collected from historical data or can be generated using the Monte Carlo sampling techniques. Therefore, for any feasible solution to the first stage problem, say $\bar{x}$, we can approximate $E[Q(\bar{x}, \xi)]$ by averaging $Q(\bar{x}, \xi_j), j = 1, 2, ..., N$ as

$$\underset{y \in \Re_+^{n2}}{Min} \quad \frac{1}{N} \sum_{j=1}^{N} Q(\bar{x}, \xi_j). \tag{1.3.3}$$

Unlike the stochastic programming approach, the Robust Optimization (RO) [7] relaxes the assumption of known probability distribution and provides optimal solutions that are feasible for a defined set of values for the uncertain parameters. In other words, RO optimizes against the worst-case realization of the random vector. The conservatism of the solution can be controlled by the means of a defined budget of uncertainty. When the problem structure allows us to divide it to a first stage problem (i.e., before the realization of uncertain parameters) and a second stage problem to adjust the solutions at the minimum cost, a two-stage robust optimization model can be developed similar to a two-stage stochastic model. The difference is in the second stage objective, where the minimum cost under the worst-case scenario from the defined uncertainty set is calculated.

The recourse-based approach to stochastic programming requires the decision-maker to assign a cost to recourse activities that are taken to ensure feasibility of the second-stage problem. In essence, the philosophy of this approach is that infeasibilities in the second stage are allowed at a certain penalty. The approach thus focuses on the minimization of expected recourse costs. In the Chance-Constrained programming (CCP) models [15], the focus is on the reliability of the system, i.e., the system's ability to meet feasibility in an uncertain environment. This reliability is expressed as a minimum requirement on the probability of satisfying constraints. A generic chance-constrained programming model can be shown as

$$\begin{aligned} \underset{x \in X}{Min} \quad & f(x), \\ & \text{subject to} \\ Pr[G(x,\xi) \leq 0] & \geq 1 - \alpha. \end{aligned} \tag{1.3.4}$$

The CCP model aims to find a solution x from the feasible set X that minimizes the function $f(x)$ while satisfying the chance constraint $G(x, \xi) \leq 0$ with a probability of at least $(1 - \alpha) \times 100\%$. It is assumed that the probability distribution of ξ is known.

The recourse-based model 1.3.1 makes a decision based on present first-stage and expected second-stage costs. In this model, the decision-maker is only interested in minimizing the average cost neglecting the variability of the outcomes for different realizations of the random vector. In other words, the decision maker is risk-neutral, i.e., does not differentiate between two solutions with the same expected value but significantly different variabilities. To capture the notion of risk in stochastic programming, a modified version of 1.3.1 can be formulated as

$$Min \quad c^T x + \lambda E_{\xi \in \Omega}[Q(x, \xi)] + (1 - \lambda)\rho(\xi, y) \tag{1.3.5}$$

where $\rho(\xi, y)$ is a risk measure that aims to minimize the variability of the outcomes under different realizations of the random parameters. The concept of using the expected value (EV) in scheduling optimization can be helpful for problems with predictable variability (i.e., low risk) on the parameter. However, for problems with frequent changes in a less predictable manner (i.e., high-risk), the optimal solution may show poor performance for specific realizations of the random data. This is attributed to the risk-neutral [74] behavior of the EV measure, which focuses on maximizing (minimizing) the expected profit (loss). This means that the EV approach treats two different solutions as equivalent if they give the same objective value without considering the characteristics of the variability such as the magnitude and/or symmetry of variance. This can be a problem if a performance measure in scheduling OR is to minimize unexpected variability to complete the planned surgeries for the day. To address this issue, we propose a solution approach based on CVaR [78] in Chapter 3, which was introduced in finance to minimize risk. CVaR has been shown to reduce the variability of performance measures, while simultaneously minimizing the expected value compared to EV [67]. Moreover, the proposed chance-constrained model in Chapter 4 achieves a moderate trade-off between minimizing total cost and the variability of costs [68].

1.4 Contributions

The proposed research aims to address surgery duration uncertainty and to make a trade-off between optimizing average performance and reducing variability of the performance measures via developing risk-based solution methods. A risk averse solution method using Conditional Value-at-Risk (CVaR) is proposed in Chapter 3 to reduce variability on overtime and its associated costs in a daily OR scheduling problem. Our numerical experiments show that the CVaR-based model proves effective in simultaneously reducing the expected value and the variance of a performance measure. The OR scheduling problem is formulated as a stochastic mixed-integer linear programming model, where a surgery duration follows a probability distribution function. Numerical results from real-life instances show that our approach not only outperforms the widely used expected value (EV) approach in reducing variability on overtimes, but it shows acceptable performance in minimizing the expected value of overtimes and idle times. We also show that the proposed method performs well under distributional uncertainty of the random data. The proposed method does not explicitly optimize the average performance but focuses on a rather small portion of scenarios. Our next work tries to tackle this limitation using chance-constrained programming.

Chapter 4 presents a chance-constrained mixed-integer programming model for the OR scheduling problem with stochastic surgery durations. The individual chance constraints control the risk of OR overtime. The goal is to minimize the sum of OR opening, OR overtime and patient waiting costs. Our model is compared with two other stochastic models in the literature: the EV model and the CVaR-based model. We demonstrate that minimizing the expected costs when solving the chance-constrained OR scheduling model results in significant savings compared to the case where only the deterministic costs are minimized. Moreover, we compare the individual and joint chance constraints in terms of allocated ORs, second-stage stochastic costs and solution times. A decomposition algorithm with strong feasibility and optimality cuts is applied to effectively solve large-scale test instances. We propose an algorithm that generates feasibility cuts using the first-stage solutions, and as a result, reduces the time required to find feasible solutions significantly. Numerical experiments demonstrates that the decomposition algorithm outperforms both the IBM CPLEX solver and a basic decomposition algorithm by solving the largest test instances to optimality within the one-hour time limit. Moreover, it is shown that the individual chance constraints lead to higher OR utilization, reduced patient waiting times and shorter solution times. The next contribution of this research focuses on developing stronger solution methods to solve very large-scale problems more efficiently.

In Chapter 5, we develop a computationally efficient decomposition algorithm augmented by strong lower and upper bounds using Lagrangian relaxation and CVaR approximation. Numerical experiments demonstrate that our decomposition algorithm outperforms the IBM CPLEX solver and the traditional Benders decomposition method in solving large-scale test instances. The proposed algorithm can solve instances with more than 200 surgeries and 40 ORs within one hour. It is also shown that the bounding methods result in finding high-quality initial feasible solutions. It is observed that the proposed algorithm finds stronger bounds for the optimal objective function compared to other methods. These bounds become tighter as the probability threshold for OR overtime increases, resulting in shorter solution times for large-scale test instances.

1.5 Outcomes

Journal Publications

- Najjarbashi, A., and Lim, G. (2015). Using augmented ϵ-constraint method for solving a multi-objective operating theater scheduling. *Procedia Manufacturing.* **3**:4448-4455.
- Lim, G. J., Mobasher, A., Bard, J. F., and Najjarbashi, A. (2016). Nurse scheduling with lunch break assignments in operating suites. *Operations Research for Health Care.* **10**:35–48.
- Najjarbashi, A., and Lim, G. J. (2019). A variability reduction method for the operating room scheduling problem under uncertainty using CVaR. *Operations Research for Health Care.* **20**:25–32.
- Najjarbashi, A., Lim, G. J. (2020). A Decomposition Algorithm for the Two-Stage Chance-Constrained Operating Room Scheduling Problem. *IEEE Access.* **8**:80160–80172.

Conference Presentations

- Najjarbashi, A. and Lim, G. (2016). Solving stochastic operating room surgery scheduling problem using conditional value-at-risk. *INFORMS Annual Conference.* Nashville, Tennessee.

- Lim, G. J., Mobasher, A., Bard, J. F., and Najjarbashi, A. (2017). A swapping heuristic for daily nurse scheduling in operating suites. *INFORMS Computing Society Conference.* Austin, Texas.

- Najjarbashi, A., and Lim, G. J. (2018). Risk-averse methods for the operating room scheduling problem under uncertainty. *INFORMS Annual Conference.* Phoenix, Arizona.

1.6 Organization

The remainder of this dissertation is organized as follows. In Chapter 2, we review the relevant literature on OR scheduling problems under uncertainty as well as risk measures, as these subjects pertain to the application area and methodological domain considered in the study. In Chapters 3, we propose our risk averse solution method to reduce variability of performance measures in the daily OR scheduling problem with surgery duration uncertainty. The proposed solution method is compared against the commonly used EV method. In Chapter 4, we describe our two-stage stochastic chance-constrained OR scheduling model and a decomposition algorithm to solve large-scale instances. In Chapter 5, we focus on developing stronger solution methods to solve even greater large-scale problems more efficiently. This chapter presents a computationally efficient decomposition algorithm augmented by strong lower and upper bounds using Lagrangian relaxation and CVaR approximation.

Chapter 2

Literature review

Researchers started to survey the operating room planning and scheduling problems since the late 70s. One of the first reviews was conducted by Magerlein and Martin [58] where they divided the problem into two major processes: 1) Advance Scheduling: scheduling patients for surgery on some future date, and 2) Allocation Scheduling: determining sequence of surgical cases on a given day. Following the work by Magerlein and Martin [58], other survey articles were published by Przasnyski [75], Smith-Daniels, Schweikhart and Smith-Daniels[86], and Blake and Carter [9] using a variety of classification techniques and frameworks. The number of published articles in the field faced a significant increase since the beginning of the 21st century. Therefore, several researchers reviewed the articles published within the past two decades. May, Spangler, Strum and Vargas [61] categorized the literature based on the planning horizon of the schedule into six groups, ranging from long-term capacity planning to the same-day scheduling and online monitoring. Similarly, Guerriero and Guido [33] categorized the literature into three main groups based on their planning horizon: strategic, tactical, and operational. Then, the articles in each group were studied in more detail using criteria such as objectives, constraints and solution approach. However, classifying the literature using planning horizon may be inaccurate since the boundaries between time frames can vary considerably for different settings. Moreover, there are several other factors in the OR scheduling problems that can be used to provide an instructive review. Samudra, Van Riet, Demeulemeester, Cardoen, Vansteenkiste and Rademakers [79] completed the work in [13] and categorized the literature using seven descriptive fields: patient characteristics, performance measures, decision delineation, upstream and downstream resources, sources of uncertainty, solution methodology, and application and experiments. They identified the trends and common selections in each field. Our research is focused on addressing surgery duration uncertainty using stochastic programming approach. Therefore, we dedicate a major portion of this chapter to review the articles using the *uncertainty* and *solution methodology* descriptive fields.

2.1 Uncertainty

One of the major problems associated with the development of accurate operating room schedules or capacity planning strategies is the uncertainty inherent to surgical services. There are several sources of uncertainty including but not limited to surgery durations, patient arrivals, delays in support services, acute onset of abnormal medical conditions (infections, chest pain, etc.) requiring delay or cancellation, inaccurate or inappropriate reservations, and lack of a mechanism to enable dynamic scheduling. Random surgery durations and emergency patient arrivals has received the most attention in the literature amongst other factors. Surgery durations are difficult to predict because for some surgeries the magnitude of the procedure only becomes apparent once the surgery is already in progress. Additionally, the durations often depend on various complex factors, e.g., the characteristics of the patient, the surgeon and the surgical team. May, Spangler, Strum and Vargas [61] elaborate on duration uncertainty by giving two main reasons for variations in the predicted surgery duration: 1) surgeons do not always know in advance all the procedures that must be performed on a scheduled patient, such as when the patient undergoes exploratory surgery or when unexpected findings occur during surgery, the time required to treat the patient may be much longer than expected, delaying subsequent patients, or much shorter than expected, possibly creating a hole in the schedule; and 2) even in cases where all procedures are known with certainty in advance, the time necessary to perform those procedures may vary significantly, due to characteristics of the operation, the surgical team, and the patient. Therefore, uncertain surgery durations can cause a large deviation from the expected completion time of all surgery cases scheduled for each day. When the deviation is significantly large, it causes an extended overtime for the surgical team to complete the scheduled cases for the day. Consequently, the hospital has to pay a substantial overtime payment, which will result in reduced revenue. Reduced OR staff job satisfaction and increased patient waiting times would be other ramifications of the uncertain completion times. On the importance of understanding and addressing uncertainty and its sources, McManus, Long, Cooper, Mandell, Berwick, Pagano and Litvak [62] stated that "Improving the healthcare system's response to variability represents an opportunity for simultaneous gains in effective capacity, cost-efficiency, improved outcomes, and patient satisfaction. A precondition to such improvement, however, is a deeper understanding of the nature and sources of variation in demand. Without such understanding and appropriate management of variability, systems such as healthcare organizations become inefficient, overwhelmed, and frustrating for all." Guerriero and Guido [33] discuss possible effects of uncertainty in planning and scheduling processes in an operating theater. Uncertainty strongly affects the time to procedure, consequently the labor cost of an OR team. For instance, a longer than predicted surgery results in a late start for the next surgery and, potentially, for the rest of the surgeries in that day's schedule. Late start results then in direct costs associated with overtime staffing when the last surgery of the day finishes later than the scheduled completion time. It is shown that mitigating the impact of disruptions in the schedule due to uncertainty can lead to higher capacity utilization and lower costs [64]. Denton, Viapiano and Vogl [23] state that while some surgeries have relatively predictable durations others may have significant variability. The combination of tight schedules and uncertainty in duration creates the need for careful consideration of OR schedules to balance the competing criteria of OR team waiting, OR idle times, and overtime. This implies that failing to address the uncertainty in durations can result in low resource utilization. Deterministic models ignore the inherent uncertainty in surgical procedures. In spite of the significance of uncertain surgery durations, there is a large body of literature that applied deterministic models to OR scheduling problems. Some of the deterministic models for the OR scheduling problem can be found in [47, 81, 89, 90]. Numerical experiments in [48] show importance of modelling the surgery duration uncertainties. Compared with a deterministic OR planning model that only considers the average emergency demand but neglects its uncertainty, the stochastic OR planning method yields about 4% reduction in overall costs. Batun, Denton, Huschka and Schaefer [5] assessed the difference between the optimal objective function value of the stochastic model against the deterministic model under uncertain surgery durations. They showed that using the stochastic approach can reduce the optimal costs up to 28% compared to the deterministic model.

2.2 Stochastic Programming

The number of papers that considered surgery duration uncertainty is almost doubled during the past decade. The majority of these papers applied variations of the stochastic programming approach to model the uncertainty in surgery duration.

Gerchak, Gupta and Henig [30] presented one of the first stochastic programming models for operating room planning problem for elective surgeries by considering uncertain surgery durations and uncertain demand for emergency surgeries. The objective function of this problem is to maximize the expected profit which is a function of random emergency case durations and random daily usage of the operating room by emergency cases. Surgery durations and OR usage by emergency cases are assumed to follow normal distributions. Authors used successive approximation method to solve the objective function recursively.

Denton and Gupta [22] proposed a two-stage stochastic linear program to minimize the expected cost of customer waiting, server idling and tardiness by considering uncertain job durations. They assumed a single-server system at which customers arrive punctually and are served in the order of their arrival (i.e., fixed sequence). They developed a sequential bounding approach based on the standard L-Shaped algorithm to solve the problem.

Guinet and Chaabane [34] dealt with the operating room scheduling problem in tactical and operational levels. First, they solved a weekly operating room scheduling problem by considering resource constraints to assign surgical cases to operating rooms. Then, a daily surgery sequencing and scheduling problem is modeled as a mixed-integer with the aim of minimizing total hospitalization and overtime costs. A heuristic method is proposed using an assignment model with resource capacity and time-window additive constraints. This method is a primal-dual Hungarian-based method where most of the constraints are integrated in the objective function.

Jebali, Alouane and Ladet [39] studied the same problem except that they considered two strategies to deal with the surgery scheduling problem: 1) modifying the assignments made in the weekly scheduling problem in order to improve resource utilization, and 2) sequencing and scheduling surgical cases without changing the assignments made in advance. In addition to surgeries at operating rooms, they take pre-operative and post-operative stages into account assuming uncertain process times at each stage. For each stage, suitable probability density functions are fit. Maximizing utilization and minimizing patients' waiting times are considered as a single objective function.

Denton, Viapiano and Vogl [23] proposed a two-stage stochastic programming model to deal with uncertain surgery durations in a single-OR surgery scheduling problem. In the first stage, sequence of surgical cases are determined while in the second stage, start times of surgical cases are set by considering random surgery durations. Three heuristic methods that sequence surgical cases based on the increasing order of mean of durations, variance of durations, and coefficient of variation of durations are proposed in order to minimize the weighted sum of the expectation of waiting time, idling time and tardiness. Obtained results from solving real data show the dominance of the second heuristic, which sequence cases in increasing order of variation of durations, in almost all test cases.

Lamiri, Xie, Dolgui and Grimaud [48] proposed a stochastic model for the operating room planning problem with elective and emergency patients. Elective surgical cases can be planned ahead while emergency cases arrive randomly and have to be performed on the day of arrival. The operating room capacity is shared among elective and emergency cases. They assumed that surgery durations for elective patients are known and deterministic.

However, total OR time required for emergency cases arriving in a specific period is a random variable. A Monte Carlo simulation combined with mixed integer programming is applied to minimize the expected OR overtime costs as well as elective patient-related costs. Following this research, Lamiri, Grimaud and Xie[49] proposed several heuristic and meta-heuristic methods to solve the problem because exact solutions for large-scale problems cannot be obtained in a reasonable amount of time using the solution method applied in their previous paper. They proposed three sequential optimization based heuristics as well as two workload based heuristics. Observation of the obtained results shows that the quality of heuristic solutions degrade as the uncertainty decrease. Simulated Annealing and Taboo Search are the meta-heuristic approaches they applied. The proposed model and methodology in these papers can be applied in hospitals who use block scheduling strategy.

Min and Yih [63] proposed a stochastic programming model to schedule elective surgeries. They assumed random surgery durations, length of stay at ICU and block capacity with known discrete distributions with finite scenario sets. The random emergency demand is implicitly incorporated in the block capacity by subtracting emergency demand from total block capacity. Capacity constraint in ICU beds is also considered in their proposed model. In order to minimize patient-related costs and overtime costs, they applied the sample average approximation (SAA) method in order to minimize the cost of patient admissions plus the expected overtime costs in the surgery blocks. They showed that the applied solution method outperforms the deterministic model using the average surgery durations and shows better convergence behavior as the sample size increases.

Denton, Miller and Balasubramanian [24] proposed two models for assignment of surgeries to operating rooms on a given day. The first model is a two-stage stochastic linear programming model with binary variables in the first stage and simple recourse in the second stage. A longest processing time-based heuristic is improved to minimize total costs including fixed costs of opening ORs and variable overtime costs by iteratively solving the problem. The second model is the robust counterpart of the stochastic model trying to minimize the maximum cost associated with an uncertainty set for surgery durations.

Begen and Queyranne [6] studied an appointment scheduling problem by considering discrete random process times given by a joint discrete probability distribution. The sequence of jobs on a single server is assumed to be predetermined. The objective is to minimize under-utilization and overtime costs. Under simple conditions on cost coefficients, they show that the objective function is submodular and L-convex and an optimal solution to the problem can be found in polynomial time. Their model can be extended to consider emergency patients and no-shows.

Gul, Denton, Fowler and Huschka [35] proposed a bi-criteria model for outpatient surgery scheduling assuming uncertain surgery durations. A log-normal probability distribution is fit to surgery durations. Optimization criteria in this paper are expected patients' waiting time and expected operating suite overtime. First, a discrete event simulation (DES) model is constructed and used to evaluate the performance of 12 heuristics by combining different sequencing and appointment time rules. The four sequencing rules applied in this paper are increasing mean of procedure time, decreasing mean of procedure time, increasing variance of procedure time, and increasing coefficient of variation of procedure time. Then, a bi-criteria genetic algorithm is applied and the obtained solutions are compared with the solutions produced by the heuristic methods. Comparisons show that expanding the computational effort with a more sophisticated GA-based method does not make significant improvement and the heuristic using the shortest processing time sequencing rule is favored over the GA-based method.

Herring and Herrmann [38] proposed a dynamic stochastic programming model for the daily surgery scheduling problem with random surgery duration. In the problem setting considered in their research, the specialties are free to choose and sequence patients within their allocated blocks as they see fit. In the meantime, specialties or surgeons that do not have allocated time submit their cases to the surgical request queue (RQ). In the period leading up to the day of surgery, OR managers try to accommodate these RQ cases by looking for unused space in rooms originally allocated to other specialties. A number of threshold-based heuristics are applied to obtain the optimal threshold policy that preserves a desired amount of operating room space for the remaining demand from the room's allocated surgical specialty.

Koeleman and Koole [45] studied an outpatient scheduling problem assuming random arrival times which follows a Poisson distribution. In addition, they considered random surgery times and possibility of no-shows by patients in their research. They assumed that a newly arrived patient must be operated right after serving the current patient undergoing surgery. A local search heuristic method is proposed to minimize a weighted sum of outpatients' waiting time and tardiness.

Mancilla and Storer [59] proposed a two-stage stochastic model for the single-OR daily surgery scheduling problem with uncertain surgery durations. The problem under study is similar to that in [23]. Sequencing and planned start times are determined in the first stage while actual start times are calculated in the second stage once surgery durations are realized. They applied the SAA method and used a benders decomposition-based heuristic algorithm to minimize expected costs associated with patients' waiting time, staff idling time and overtime. This heuristic algorithm determines the sequence of surgeries (master problem) and then performs the scheduling (subproblems).

Bruni, Beraldi and Conforti [10] modeled the problem using three recourse strategies in order to model different reactive scheduling policies under a block booking system. The recourse strategies are overtime recourse, swapping recourse and complete rescheduling. Surgery durations and arrival of emergency patients are assumed random in a weekly planning horizon. Operating room capacity is the only resource constraint taken into account and the objective is to maximize total revenue from operating surgical cases with different priorities. An improvement heuristic method is proposed to solve the numerical instances.

Freeman, Melouk and Mittenthal [29] proposed a MIP model for the daily OR scheduling problem with stochastic surgery times and random emergency arrivals. They used break-in-moments to accommodate the emergency operations throughout the day. A two-stage heuristic solution method is developed and lower bounds are generated by reducing the problem to a multiple knapsack problem. They compared the performance of the proposed scenario-based model against deterministic and job hedging methods using two sets of methodological and comparative experiments. They showed that the proposed model improves profit and OR utilization while reducing the average waiting times for the emergency patients.

Xiao, Van Jaarsveld, Dong and Van De Klundert [96] developed a three-stage model for the single-OR scheduling problem with uncertain surgery times and resource availability. The first stage deals with the sequencing and scheduling decisions. The patients are assigned to two shifts where the patients in the first shifts must be operated. The second stage decides if the second-shift patients remain on today's schedule or moved to the waiting list for the following days. In the final stage, second shift durations become known and proper recourse decisions are made. They used the SAA approach to model the problem and applied an L-shaped algorithm to solve instances derived from a hospital in China.

Pang, Xie, Song and Luo [73] developed a stochastic integer programming model for the OR scheduling problem with surgery duration and case cancellation uncertainty. The proposed model attempted to minimize patient- and hospital-related costs. The Benders decomposition algorithm was used to solve numerical examples. They showed that the proposed model reduces total cost by 27% using a case study based on two departments at West China Hospital.

Wang, Zhang, Zhang, Tang and Mu [93] proposed a stochastic programming model for the integrated OR and surgeon scheduling problem with the objective of minimizing operating room staffing costs. A patient preference-driven policy is proposed to satisfy patients' personalised preferences for surgeons and surgery dates for high-end private hospitals. They developed a column generation-based heuristic algorithm to solve the stochastic model. The performance of the algorithm is tested on different scale instances. They showed that the proposed heuristic method can obtain solutions within a 1.6% gap of the lower bound obtained by the LP relaxation of the MIP model.

M'Hallah and Visintin [36] proposed a stochastic model for scheduling surgeries during a cyclic master surgical schedule (MSS). The model determines the number and type of surgeries to schedule in a two-week planning horizon where each operating session is assigned to a surgical specialty according to a fixed grid. They assumed stochastic surgery duration, intensive care unit time and post-surgery length of stays and accounted for the availability of both intensive care unit beds and post-surgery beds. The goal is to maximize the expected operating room throughput. They applied the sample average approximation method to approximate the stochastic parameters in the model. The effectiveness of the proposed model is studied on a case study from a European Children's Hospital.

Robust optimization is another technique that is applied to surgery scheduling problems under uncertainty. Rath, Rajaram and Mahajan[77] developed a two-stage stochastic model for the integrated anesthesiologist and OR scheduling problem. In the first stage, surgical cases are assigned to available ORs and anesthesiologists aiming to minimize the total OR opening and on-call anesthesiologist assignment costs. In the second stage, they applied the robust optimization approach to minimize the maximum overtime costs given the pre-specified budget of uncertainty. They used a data-driven approach to estimate the uncertainty sets for the random surgery durations.

Neyshabouri and Berg [70] proposed a two-stage robust model for the surgery planning problem under uncertain surgery durations and length of stay in ICU. The ICU beds are considered as limited downstream resources. It is not known until after the surgery whether a patient needs to go to the ICU. The objective is to minimize the patient admission costs, patient cancellation costs, surgery block overtime costs, and ICU capacity violation costs. A column-and-constraint generation algorithm is applied to solve the two-stage model.

2.3 Risk-Based Approach

The recourse-based approaches reviewed in the previous section requires the decision-maker to assign a cost to recourse activities that are taken to ensure feasibility of the second-stage problem. In essence, the philosophy of this approach is that infeasibilities in the second stage are allowed at a certain penalty. The approach thus focuses on the minimization of expected recourse costs using some approximation methods such as the widely used SAA method. In the chance-constraint approach, the focus is on the reliability of the system, i.e., the system's ability to meet feasibility in an uncertain environment.

Shylo, Prokopyev and Schaefer [84] proposed a stochastic model for operating room planning problem under uncertain surgery durations. Considering a block booking policy, the objective is to minimize the expected OR block idle times. They used individual chance-constraints for every OR block to ensure that the risk of overtime more than a pre-specified level is under control. Cumulative surgery durations are approximated using a normal distribution. This approximation is only appropriate for those surgical specialties with high demand level. Using the properties of overtime and under-time functions, a mixed-integer model is formulated that provides lower and upper bounds on the optimal solution of the stochastic model. A batch scheduling algorithm is proposed that solves the problem iteratively and tries to improve the lower bound on the optimal solution and provide near-optimal solution to the stochastic problem.

Zhang, Denton and Xie [98] studied the OR surgery allocation problem with uncertain surgery durations. They proposed two different chance constrained models based on assumptions on the available distributional information. First, they assumed known probability distributions for the random parameters and modelled the individual chance constraints to ensure preventing OR overtime. Second, they assumed insufficient information on the random surgery durations and formulated a distributionally robust chance constrained model. They applied a column generation reformulation of the problem to divide the original model into a master problem and subproblems. Then, they developed a branch-and-price algorithm to solve the reformulated model. They also provided lower bounds and upper bounds on the subproblems using conditional value-at-risk and probabilistic covers, respectively.

Deng, Shen and Denton [18] proposed two chance-constrained programming models for the surgery planning problem with stochastic surgery durations. In the first model, it is assumed that distributional information is available for the random parameters. On the other hand, the second model is a distributionally robust chance constrained model assuming ambiguous distributions for the random surgery durations. Two types of chance constraints are formulated to ensure the specified risk tolerances on patient waiting times and OR overtime. The distributionally robust model constructs confidence sets of all the possible distributions that results in a chance constrained model with higher reliability requirements for both waiting time and overtime. The authors developed a decomposition-based cutting plane algorithm to solve the models in order to minimize the OR opening costs as well as meeting the required quality of service levels.

Deng and Shen [19] studied the OR scheduling problem under stochastic surgery duration with a joint chance constraint for limiting the risk of OR overtime. They developed a two-stage stochastic model in a way that the second stage can be decomposed and solved separately for each OR. Each surgery must be operated in a pre-specified time window and the goal is to minimize the total OR usage costs. They proposed bounds on the proposed model and applied cutting-plane approaches to improve the computational efficiency. A brand-and-cut algorithm is developed to solve the two-stage model. A relaxed master problem as a chance constrained binary packing problem is used in order to provide better solutions to the second stage problem.

Jebali and Diabat [40] considered a surgery planning problem on a weekly planning horizon by assuming limited capacity of the ICU beds. Three sources of uncertainty are assumed in their research: surgery duration, length of stay in ICU, and the arrival of emergency patients that is incorporated as random ICU capacity. They developed a two-stage stochastic programming model with chance constraints on the violation of ICU capacity. The aim of the model is to minimize patient-related costs and the expected penalty costs for violating the ICU beds capacity. They applied the SAA method to solve the proposed model and showed that the objective value converges to that of the true problem as the sample size grows. Noorizadegan and Seifi [71] proposed a chance-constrained programming model for the surgery planning problem with uncertain surgery durations. The individual chance constraints ensure that the violation of OR time limits do not exceed the specified risk tolerances. Surgeon assignments to the ORs are also considered in their model. They modeled the uncertain surgery durations in two ways: 1) a set of discrete scenarios (e.g., historical data) are available, and 2) the probability distributions are known. A column generation formulation is developed for the problem where the master problem has a set-partitioning problem structure and the subproblems can be decomposed over ORs and time periods. The subproblems are reformulated as shortest path problems to accelerate the solution process and solved using a search algorithm. They showed that using a continuous approximation of distribution functions lead to reduction of the expected probability of exceeding OR overtime compared to discrete scenarios. They also found that the performance of the proposed method depends heavily on the search space of the column generation subproblem and efficiency of the employed branch-and-bound method.

Wang, Li and Peng [92] proposed a distributionally robust chance-constrained model for the surgery planning problem with stochastic surgery durations. It is assumed that the first two moments (i.e., mean and covariance) for the random parameters are known. The chance constraints control the risk of facing OR overtime within the specified limits. They showed that the model can be reformulated as a second order cone programming problem. They considered the ICU and ward beds as limited downstream resources and tried to minimize their peak demand during the planning horizon as well as operation-related costs.

Kamran, Karimi and Dellaert [41] proposed a two-stage stochastic programming and a two-stage chance-constrained stochastic programming for the advance scheduling problem with surgery duration uncertainty. The proposed models attempt to minimize several criteria such as patients waiting time, tardiness, cancellation, block overtime, and the number of surgery days of each surgeon within the planning horizon. The stochastic models are approximated using the SAA method and the Benders decomposition algorithm is applied to solve real-life numerical examples.

Deng, Shen and Denton [21] developed a distributionally robust chance-constrained programming model for the surgery planning problem. The proposed model make OR opening, surgery allocation, sequencing and start time decisions. It is assumed that the lack of historical data leads to unknown distributional information. They used ϕ-divergence measures to build an ambiguity set of possible distributions of random surgery durations, and derived a branch-and-cut algorithm for optimizing a mixed-integer linear programming reformulation based on finite samples of the random surgery durations. The proposed methodology is tested on real hospital-based surgery data.

2.4 Identified Challenges

We showed in the previous sections that the stochastic programming models have been widely used in the literature to address the OR scheduling problems with uncertain surgery durations. The majority of these articles attempted to minimize the expected value of penalty costs corresponding to some performance measure (e.g., waiting time, overtime) after the random parameters are realized. The concept of using the expected value (EV) in scheduling optimization can be helpful for problems with predictable variability (i.e., low risk) on the random parameter. However, for problems with frequent changes in a less predictable manner and in short-term (i.e., high-risk), the optimal solution may show poor performance for specific realizations of the random data. This is attributed to the risk-neutral behavior of the EV measure, which focuses on maximizing (minimizing) the expected profit (loss). In other words, optimizing the EV ignores the variability of the performance measure in different scenarios. This means that the EV approach treats two different solutions as equivalent if they give the same objective value without considering the characteristics of the variability such as the magnitude and/or symmetry of variance. This can be a problem if a performance measure in scheduling OR is to minimize unexpected variability to complete the planned surgeries for the day. In order to enhance predictability of an operating room (OR) schedule by minimizing variability of an outcome measure in the worst case, Chapter 3 proposes a risk-based OR scheduling model that pursues reducing variability of costs associated to overtime and idles time as well as the expected value of the costs. The OR scheduling problem is formulated as a stochastic programming model, where a surgery duration follows a known probability distribution function. To the best of our knowledge, this is the first study that proposes a CVaR-based optimization model for the OR scheduling problem underuncertainty. We use real-life test instances collected from a Dutch hospital [50] and use several criteria to compare the CVaR-based model with the EV model. Table 2.1 summarizes the selected published literature on chance-constrained programming (CCP) models and identifies the research gaps that are addressed in this chapter. First, it can be observed that very few articles proposed a CCP model for the OR scheduling problem under uncertainty [19, 21]. Surgery scheduling problems often have a more complex structure resulting from a variety of decisions, such as OR opening, patient-to-OR assignment, surgery sequencing, and projected and actual start times before and after the realization of random surgery durations, respectively. CCP-based models have the potential to effectively handle such large variabilities in daily surgery scheduling problems [61].

Second, a majority of the models have neglected the importance of minimizing the stochastic second-stage costs. Their primary focus has been on providing schedules within the specified risk tolerances while also aiming to minimize deterministic performance measures, such as fixed OR opening costs [21, 71]. Unlike existing approaches, Chapter 4 proposes a chance-constrained model that aims to minimize both deterministic and stochastic costs for the OR scheduling problem. The significance of considering both classes of costs is highlighted using numerical experiments.

Third, we provide insightful observations about the performance of three different models (CCP, CVaR and EV) in solving the stochastic OR scheduling problem under various risk thresholds. The proposed model is compared alongside EV and CVaR models using several metrics such as total costs, OR utilization and solution time. Moreover, the performances of both individual and joint chance constraints are compared in terms of OR opening decisions, minimizing the second-stage costs and computational efficiency.

Finally a computationally efficient decomposition algorithm is applied to provide high-quality solutions for the large-scale test instances within reasonable time frames. We proposed an algorithm to derive feasibility cuts using the first-stage solutions that accelerate finding feasible solutions and the convergence speed.

Table 2.1: Research Gaps in the Chance-Constrained OR Scheduling Literature

Authors (Year)	**Decisions**			**Objective**	
	OR	**Resource Allocation**	**Sequencing & Scheduling**	**Deterministic Costs**	**Stochastic Costs**
Shylo et al. (2012)		✓			OR Idle Time
Zhang et al. (2015)		✓			OR Overtime
Deng et al. (2019)	✓	✓	✓	OR Opening	
Noorizadegan & Seifi (2018)	✓	✓	✓	OR Opening & Turn-Over	
Wang et al. (2017)	✓	✓		Operational	
This Work	✓	✓	✓	OR Opening	OR Overtime & Patient Waiting Time

The stochastic programming models are often note scalable to large models. Hence, a large body of heuristic and metaheuristic methods in the literature aim to solve the optimization fast, but they do not guarantee optimality [65, 91, 95]. Decomposition algorithms have been widely used in the literature to solve large-scale mixed-integer programming models [76]. Lagrangian methods were applied in the early 1970s to general integer programming problem [28, 82] and scheduling problems [26]. Several papers in the literature have used Lagrangian relaxation to solve the OR scheduling problem [3, 32, 99]. In Chapter 5, we develop a computationally efficient decomposition algorithm that is augmented by strong lower and upper bounds using Lagrangian relaxation and CVaR approximation. We show that the proposed algorithm outperforms the well-known Benders decomposition method and achieves the global optimal solution for large-scale test instances in less than 60 minutes. Moreover, the proposed lower and upper bounds obtain high-quality initial feasible solutions and perform consistently for different probability thresholds.

Chapter 3

A Variability Reduction Method for the Operating Room Scheduling Problem under Uncertainty using CVaR

3.1 Introduction

An operating suite refers to a medical complex which includes multiple rooms designed and equipped to perform surgeries. Operating suites contribute to more than 30% of total expenses [56] and 40% of total revenues [23] in hospitals. One of main challenges dealing with operating suites is unexpected surgical interventions occurring every day [61]. This uncertainty hampers the efficiency of the provided schedules along with incurring excessive costs to hospitals due to improper resource utilization in practice. In order to enhance predictability of an operating room (OR) schedule by minimizing variability of an outcome measure in the worst case, a risk-based OR scheduling model is proposed in this chapter. The objective of the model is to reduce variability of a performance measure (e.g., overtime and/or idles time associated expenses) as well as the expected value of the measure.

Cardoen, Demeulemeester and Beliën [12] categorizes the operating room scheduling process into four planning levels: case mix planning (CMP), master surgery scheduling (MSS), daily OR scheduling, and adaptive scheduling. The daily OR scheduling problem itself is divided into two steps. The first step is a longer-term planning in which patients are scheduled for certain days over a given planning horizon (e.g., weekly, monthly). The second step is to determine the sequence of surgery cases with surgery start time for the specific day. This work focuses on the second step of the daily OR scheduling problem for a set of elective [12] patients, whose surgery durations follow a known probability distribution function (PDF).

Several good survey articles in OR planning and scheduling have been published over the years [9, 13, 58, 75]. Samudra, Van Riet, Demeulemeester, Cardoen, Vansteenkiste and Rademakers [79] classified the literature using seven different perspectives of OR scheduling problems. May, Spangler, Strum and Vargas [61] found high variability of surgery durations as the biggest challenge towards developing practical OR schedules among many sources of uncertainty mentioned in their review. The required time for surgical interventions may vary significantly based on the type of operations being performed, the surgical team and the patient. Due to such uncertainty in scheduling, reducing variability of the performance measures in the provided schedules can help elevate capacity utilization, cost-efficiency and patient satisfaction [62]. Hence, this study aims to develop an OR scheduling approach that can effectively tackle the uncertain surgery times and control the variability of the performance measure.

Stochastic programming models have been widely used in the literature to address the OR scheduling problems with uncertain surgery durations [10, 20, 23, 40, 48, 59, 63, 64, 77, 84]. The majority of these articles attempted to optimize the expected value of a performance measure. The concept of using the expected value (EV) in scheduling optimization can be helpful for problems with predictable variability (i.e., low risk) on the parameter. However, for problems with frequent changes in a less predictable manner and in short-term (i.e., high-risk), the optimal solution may show poor performance for specific realizations of the random data [67, 72]. This is attributed to the risk-neutral [74] behavior of the EV measure, which focuses on maximizing (minimizing) the expected profit (loss). Consequently, the EV approach treats two different solutions as equivalent if they give the same objective value without considering the characteristics of the variability such as the magnitude and/or symmetry of variance. This can be problematic if a performance measure in scheduling OR is to reduce unexpected variability to complete the planned surgeries for the day.

This could be achieved by minimizing a direct measure such as variance [60]. However, using variance as the risk measure has a number of drawbacks. First, calculating the variance involves quadratic expressions that result in a non-linear discrete optimization model. Solving such models for large-scale instances can be computationally difficult. Second, variance shows a poor performance when the probability distribution of the random parameters are non-symmetric [43] because the skewness of the random parameters, if exists, is not fully reflected. The variation in the daily OR schedule is identified as one of the main factors that constrain OR productivity and efficiency. Cima, Brown, Hebl, Moore, Rogers, Kollengode, Amstutz, Weisbrod, Narr, Deschamps and Team S.P.I [17] emphasized designing surgical scheduling processes that reduce the variation of both under- and over-utilization of OR resources. Smith, Spackman, Brommer, Stewart, Vizzini, Frye and Rupp [85] showed that a 20% decrease in the daily schedule variation can reduce the staff turnover rate by 41%. CVaR has been shown to reduce variability of performance measures compared to EV[67], while simultaneously reducing the expected value [80].

Therefore, this chapter presents a risk-based solution approach using the concept of Conditional Value-at-Risk [78] to reduce variability on overtime, idle time, and associated costs in a daily OR scheduling problem. The OR scheduling problem is formulated as a stochastic mixed-integer linear programming (SMILP) model, where a surgery duration follows a probability distribution function. The objective of the SMILP model is to minimize the CVaR of overtime and idle time costs. CVaR [78] was introduced in finance to minimize the extreme losses in the tail of the distribution of possible return. An advantage of using a CVaR approach is that the resulting model can be reduced to a linear programming model. CVaR takes the skewness of the random parameters into account. To the best of our knowledge, this is the first study that proposes a CVaR-based optimization model for the OR scheduling problem under uncertainty.

The remainder of this chapter is organized as follows. Section 3.2 describes the daily

OR scheduling problem considered in this work. In Section 3.3, the CVaR measure is first discussed, and the proposed stochastic MILP model using CVaR is described for the OR scheduling problem. Numerical experiments are conducted on a benchmark set in Section 3.4 and the observations are reported. Conclusions are drawn in Section 3.5.

3.2 Problem Description

We consider operating suites with R operating rooms. For a given set (I) of elective patients each day, the problem is to determine the assignment of patients to available OR rooms, and the sequence of surgeries to be performed so as to minimize the total cost associated with excessive overtimes and idle times. We assume that our operating suites use the block booking strategy [25] with B blocks. Each OR block is to be assigned to a surgical team or a surgeon via the master surgery schedule. An incidence matrix $E_{I\times R}$ is constructed to show the eligible patient-to-OR assignments. The surgery durations are uncertain with a known probability distribution function. The Monte Carlo sampling method is used to produce S scenarios comprising the realizations of the stochastic surgery durations.

There are three decision variables in the OR scheduling problem considered in this research, namely: (1) patient-to-room assignment, (2) surgery sequencing, and (3) surgery start time. In *surgery sequencing*, an overlap is not allowed between two successive operations because operations cannot be interrupted or stopped once they are started. The goal is to minimize the sum of overtime and idle time costs. In the remainder of this chapter, we refer to the performance measure as the *total cost*. The OR cost is divided into overhead cost and variable cost as in [57]. The overhead cost is the same among all ORs, whereas the variable cost can differ from one OR to another depending on surgeon-to-OR assignments from the MSS. Section 3.4.1 provides more detail on parameter settings. The following section introduces CVaR and explains how we utilize it to develop a risk-based optimization model for the daily OR scheduling problem under uncertainty.

3.3 Risk-based OR Scheduling

3.3.1 Conditional Value at Risk: a brief introduction

Let $f(x, y)$ be the objective function with the decision vector x and the random vector y having density $p(y)$. Then, for a fixed x, the objective function is a random variable. Given a probability level $\alpha \in (0, 1)$, the corresponding Value-at-Risk (VaR) and the CVaR of the objective function can be formulated, respectively, as

$$\zeta_\alpha(x) = Min\,\{\zeta \in Re : \psi(x, \zeta) \geq \alpha\} \tag{3.3.1}$$

$$\text{and} \qquad \phi_\alpha(x) = \frac{1}{1-\alpha} \int_{f(x,y) \geq \zeta_\alpha(x)} f(x, y)p(y)dy \tag{3.3.2}$$

where $\psi(x, \zeta)$ is the cumulative density function (CDF) of $f(x, y)$ for fixed x. Therefore, $\phi_\alpha(x)$ is the expectation of those outcomes of $f(x, y)$ that exceeds VaR. The CVaR function $\phi_\alpha(x)$ can be simplified as [88]

$$F_\beta(x, \zeta) = \zeta + \frac{1}{1-\alpha} \int_{y \in Re^m} [f(x, y) - \zeta]^+ p(y)dy \tag{3.3.3}$$

where $t^+ = max\{t, 0\}$. $F_\beta(x, \zeta)$ is convex with respect to ζ, and minimizing it gives the minimum CVaR and calculates VaR, simultaneously. By generating a finite number of scenarios from the density function of the random vector y, $F_\beta(x, \zeta)$ can be approximated as

$$\tilde{F}_\beta(x, \zeta) = \zeta + \frac{1}{s(1-\alpha)} \sum_{k=1}^{s} [f(x, y_k) - \zeta]^+ p(y)dy \tag{3.3.4}$$

where scenarios are assumed to be equally likely. When $f(x, y)$ is linear with respect to x, minimizing $\tilde{F}_\beta(x, \zeta)$ can be reduced to a linear programming (LP) problem [88].

Artzner, Delbaen, Eber and Heath [2] stated a set of properties that should be desirable for a risk measure. Any risk measure which satisfies these axioms is said to be *coherent*. The four axioms are monotonicity, translation equivariance, subadditivity, and positive homogeneity. CVaR is a coherent risk measure for all $p(y)$ whereas VaR is not coherent when the random data follow non-Normal distributions. VaR is difficult to optimize when it is approximated using finite scenarios and it may show multiple local extrema. Moreover, optimizing CVaR gives the VaR at the same confidence level as a by-product.

In the next section, we present the MILP formulation to minimize the CVaR of total cost for the OR scheduling problem with stochastic surgery durations.

3.3.2 Mathematical Formulation

In practice, historical data and knowledge about surgery cases are commonly used to estimate elective surgery durations [14, 29]. This study assumes that a surgery duration follows a known probability distribution function.

This makes the optimization model extremely difficult to solve due to (1) the corresponding model having a large number of integer variables, and (2) the presence of the stochastic variables (i.e., surgery duration). Therefore, a scenario-based formulation is proposed to address the computational burden to solve the model [29]. Our MILP model formulation is based on a finite set of scenarios generated using the Monte Carlo sampling method, where each scenario contains a specific realization of the elective surgery duration. Using this information, the objective is to minimize the costs associated to overtimes and idle times. We refer to the resulting formulation as a scenario-based daily OR scheduling (SDORS) model. Consider the following notation for the SDORS model:

Sets:

I set of elective patients $I = \{1,...,\text{I}\}$

R set of operating rooms R=$\{1,...,\text{R}\}$

S set of scenarios S=$\{1,...,\text{S}\}$

Indices:

i, j, k elective patients $i, j, k \in I$

r operating rooms $r \in R$

s scenarios $s \in S$

Parameters:

E_{ir} incidence matrix for eligible patient-to-OR assignments

d_{is} surgery duration for patient i under scenario s

μ_i mean value of surgery duration for patient i

c_r expected open time of operating room r

w_r operation cost per minute for the surgeon working in room r during regular hours

π overtime cost coefficient, $\pi > 1$

α probability level, $\alpha \in (0, 1)$

M a sufficiently large number

Variables:

$$x_{ir} = \begin{cases} 1 & \text{if patient } i \text{ is assigned to room } r \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ijr} = \begin{cases} 1 & \text{if patient } i \text{ precedes patient } j \text{ in room } r \\ 0 & \text{otherwise} \end{cases}$$

t_{is} surgery start time for patient i under scenario s

ov_{rs} overtime of room r under scenario s

id_{rs} idle time of room r under scenario s

ζ VaR of total overtime and idle time costs

z^s amount of cost exceeding the threshold value *zeta* under scenario s

The formulation of the SDORS model is given as follows:

$$Min \left[\zeta + \frac{1}{|S|(1-\alpha)} \sum_{s \in S} z^s\right], \tag{3.3.5}$$

$$\zeta + z^s \geq \sum_r \pi w_r ov_{rs} + \sum_r w_r id_{rs} \qquad \forall s \in S, \tag{3.3.6}$$

$$ov_{rs} \geq \sum_{i \in I} x_{ir} d_{is} - c_r \qquad \forall r \in R,\ s \in S, \tag{3.3.7}$$

$$id_{rs} \geq c_r - \sum_{i \in I} x_{ir} d_{is} \qquad \forall r \in R,\ s \in S, \tag{3.3.8}$$

$$\sum_{r \in R} x_{ir} = 1 \qquad \forall i \in I, \tag{3.3.9}$$

$$x_{ir} \leq E_{ir} \qquad \forall i \in I,\ r \in R, \tag{3.3.10}$$

$$\sum_{i \in I} \mu_i x_{ir} \leq c_r \qquad \forall r \in R, \tag{3.3.11}$$

$$y_{ijr} + y_{jir} \leq \frac{1}{2}(x_{ir} + x_{jr}) \qquad \forall i \neq j \in I,\ r \in R, \tag{3.3.12}$$

$$y_{ijr} + y_{jir} \geq x_{ir} + x_{jr} - 1 \qquad \forall i \neq j \in I,\ r \in R, \tag{3.3.13}$$

$$y_{ijr} + y_{jkr} - 1 \leq y_{ikr} \qquad \forall i \neq j \neq k \in I,\ r \in R, \tag{3.3.14}$$

$$t_{is} + d_{is} - t_{js} \leq M(1 - y_{ijr}) \qquad \forall i \neq j \in I, r \in R, s \in S, \tag{3.3.15}$$

$$z^s \geq 0 \qquad \forall s \in S, \tag{3.3.16}$$

$$\zeta \quad unrestricted, \tag{3.3.17}$$

$$x_{ir}, y_{ijr} \in \{0, 1\} \qquad \forall i \neq j \in I, r \in R, \tag{3.3.18}$$

$$ov_{rs}, id_{rs} \geq 0 \qquad \forall r \in R, s \in S, \tag{3.3.19}$$

$$\text{and} \quad t_{is} \geq 0 \qquad \forall i \in I, s \in S. \tag{3.3.20}$$

The objective function (3.3.5) minimizes the CVaR of total cost associated with overtimes and idle times. For each scenario, constraint (3.3.6) along with constraints (3.3.16 – 3.3.17) determines z^s as the amount of cost that exceeds, if at all, the threshold value of

ζ. Constraints (3.3.7,3.3.8) calculate the overtime and idle time for each OR in each scenario, respectively. Constraint (3.3.9) ensures that all patients will be operated on during the day. It is assumed that each OR is equipped for specific surgical specialties. Therefore, Constraint (3.3.10) enforces eligible patient-to-OR assignments. Constraint (3.3.11) ensures that the sum of the mean duration of the assigned patients to an OR does not exceed its expected open time. Constraints (3.3.12-3.3.14) enforce the sequencing rules. Constraint (3.3.12) prevents sequencing of cases that are not assigned to the same OR. Constraint (3.3.13) along with constraint (3.3.12) enforces surgery i to start before/after surgery j if they are allocated to the same room. Constraint (3.3.14) defines the order transitivity of a schedule such that surgery i starts before surgery k, if surgery i starts before j and surgery j precedes surgery k, for any distinct surgeries $i, j, k \in I$. Constraint (3.3.15) determines the operation start times according to the sequencing decisions. The proposed MILP model with the CVaR objective function is called the *SDORS-CVaR* model. In a special case, CVaR measure can be converted to EV by setting the confidence level (α) to zero. The resulting model is called *SDORS-EV* and will be compared to SDORS-CVaR later in Section 3.4.

3.3.3 Strengthening the big M parameter

A proper value for the big M parameter can be calculated to tighten the feasible region of the model. Assuming that $y_{ijr} = 0$, the left-hand-side of constraint (3.3.15) takes the largest value when $t_j = 0$ and case i is the last surgery scheduled in room r. Therefore, a sufficiently large M can be calculated as

$$M = \underset{r \in R, s \in S}{Max} \left\{ \sum_i E_{ir} d_{is} \right\}. \tag{3.3.21}$$

3.4 Numerical Experiments

The proposed SDORS-CVaR model is evaluated using a set of eight numerical examples in this section. The optimal schedules obtained by SDORS-CVaR are compared to that of the SDORS-EV model (SDORS-EV) in terms of the total cost associated with the schedule. The comparison is based on the ability of the models in reducing the variability on total cost as well as minimizing the expected value of total cost. Section 3.4.1 explains how the test problems are set up for experiments, followed by an analysis of the results (Section 3.4.2) and performance evaluations on variability reduction of the proposed method (Section 3.4.3). Using robust and non-robust measures of variability [37], it is observed that the SDORS-CVaR model outperforms the SDORS-EV model in reducing the variability of the performance measure. The SDORS-CVaR also outperforms the SDORS-EV in reducing the worst-case cost. Moreover, the SDORS-CVaR model offers a great level of flexibility to the decision-maker to incorporate his or her risk preferences into the created schedule by changing the probability level α.

3.4.1 Experiment Setup

A set of eight test problem instances are obtained from [50]. The instances are different from each other in terms of surgery types and specialties such as neurology, orthopedic, thoracic, and oncology. The surgery durations are assumed to follow a three-parameter log-normal distribution [29, 35, 87]. The instances are labeled in numeric order $\{1, 2, \cdots, 8\}$, and they vary greatly in problem size, i.e., number of patients and number of ORs. These cases are selected to appropriately test the ability of the proposed model in handling problems with different complexity levels.

There are a total of 432 surgical cases in all eight instances. The mean surgery duration is between 1-2 hours in almost half of the cases (197 out of 432). In instances {1,2,4,5,6}, 45% of the cases (93 out of 208) require more than two hours of OR time, while 46% of the cases (104 out of 224) in instances {3,7,8} take less than one hour on average. All instances have low to moderate variability based on their coefficient of variation (CV), that ranges between 55% and 71%. We used the estimates of the OR overhead ($10 per minute) and variable costs presented in [55, 57]. However, choosing a proper value of the OR variable cost is not straightforward due to large variations among different surgeons. In this study, each OR has one block that is assigned to a surgeon to perform surgical cases. Hence, the OR variable cost varies from one surgery case to the next depending on the surgical team. For example, according to 2,848 surgery cases reported in [55], the average variable cost ranged between \$9 and \$34 per minute.

All numerical experiments in this chapter are based on the following configuration. One hundred scenarios are generated for random surgery duration using the Monte Carlo sampling method. All ORs are planned to operate for 9 hours [29] for the day. The sum of the expected surgery durations in each OR should not exceed its availability. However, if a surgery duration in a generated scenario exceeds the limit for the OR availability, an overtime cost will incur according to the amount of violation. Optimization models are implemented in GAMS modeling language and solved using the IBM CPLEX 12.6 solver on a workstation with 24 cores, 3 GHz processors, and 384 GB of memory. Table 3.1 summarizes the data used in our numerical experiments.

Table 3.1: Numerical instances and parameter settings

Instance ID	1	2	3	4	5	6	7	8
Number of Patients	9	22	34	40	63	74	89	101
Number of ORs	5	5	5	10	20	20	15	20
Average CV	0.667	0.628	0.579	0.617	0.551	0.713	0.605	0.585
PDF	Three-parameter Log-normal							
OR availability (minute)	540							
OR blocks	1							
Overhead cost (\$/minute)	10							
Variable cost (\$/minute)	Ranges between 9 and 34							
Scenarios	100							

3.4.2 Numerical Results

In this section, we first show that using the SDORS-CVaR model reduces the variation of the sum of overtimes and idle times in the daily schedule. Once the SDORS model generates a schedule for a given problem instance, the sum of overtimes and idle times, and the corresponding total cost are calculated for all scenarios ($|S|$). As a result, a vector $\tau \in \mathbb{R}^{|S|}$ is generated, in which each entry corresponds to the sum of overtime and idle time of a scenario. We call τ a *utilization vector.* Each problem instance is solved using both SDORS-CVaR and SDORS-EV models, and the corresponding utilization vectors are obtained. Note that the SDORS-EV model is equivalent to the SDORS-CVaR when the confidence level parameter (α) equals zero. Then, the average and the standard deviation of the utilization vector are calculated.

We define two parameters $\mu_\Delta(\alpha)$ and $\sigma_\Delta(\alpha)$ in a normalized scale to make performance comparison between the two models. Let $\mu_\Delta(\alpha)$ be the percentage of difference in the average value between the two models (CVaR - EV) and it is calculated based on the average values $\mu_\aleph(\alpha)$ of the total cost corresponding to model $\aleph \in \{CVaR, EV\}$

$$\mu_\Delta(\alpha) = \frac{\mu_{CVaR}(\alpha) - \mu_{EV}(\alpha)}{Max\left\{\mu_{CVaR}(\alpha), \mu_{EV}(\alpha)\right\}} \times 100\%.$$

Similarly, let $\sigma_\Delta(\alpha)$ be the percentage of difference in standard deviations, $\sigma_\aleph$

$$\sigma_\Delta(\alpha) = \frac{\sigma_{CVaR}(\alpha) - \sigma_{EV}(\alpha)}{Max\left\{\sigma_{CVaR}(\alpha), \sigma_{EV}(\alpha)\right\}} \times 100\%.$$

Figure 3.1 shows these comparisons where the *x-axis* corresponds to its average value difference $\mu_\Delta(\alpha)$ and the *y-axis* is for $\sigma_\Delta(\alpha)$. As the value of α increases there shows a clear advantage of using the CVaR in reducing cost variability as compared to the EV, i.e., the value of $\sigma_\Delta(\alpha)$ was decreased by almost 18%. However, an opposite trend was observed on the average cost as it continues to climb up as the value of α increases. Hence, one must be careful about choosing an appropriate value of α that works well for the organization's goal.

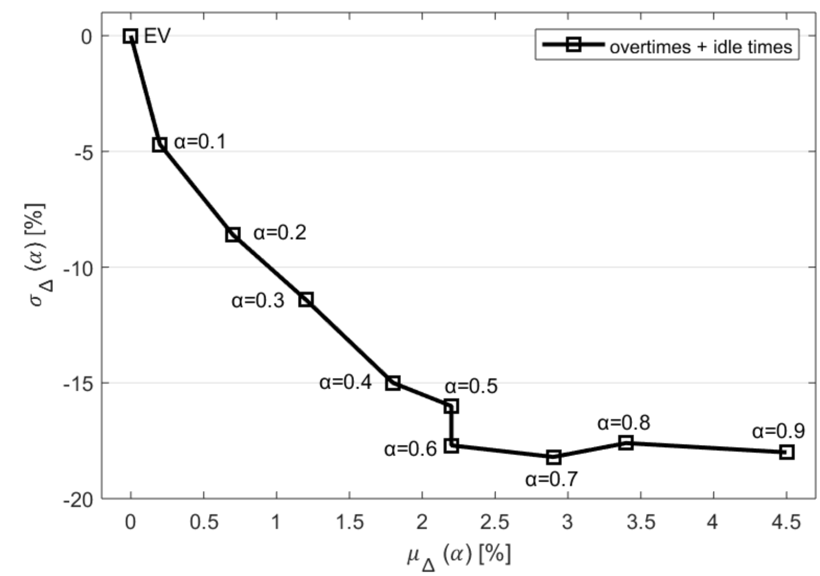


Figure 3.1: CVaR vs. EV in reducing the variance of overtimes and idle times

We now analyze the sensitivity of the CVaR model on $\mu_{CVaR}(\alpha)$ and $\sigma_{CVaR}(\alpha)$ of the total cost as we vary the value of α. Figure 3.2 shows the correlation between the average value and the standard deviation of the *total cost (overtime cost + idle cost).* As we increase the value of α, the total cost variability reduces and the average cost increases. For example, using the SDORS-CVaR with $\alpha = 50\%$ reduces the standard deviation of total cost from \$5,435 to \$3,968 compared to the SDORS-EV at the price of increasing the total cost from \$12,680 to \$14,000. Hence, a decision-maker should look at such trade-offs between the decrease of variability and the corresponding increase of the average cost and select an appropriate schedule.

Further numerical results are analyzed to show the advantages of using the CVaR over the EV. The cost vector , $\chi \in R^{|S|}$, is used for this purpose, in which each entry of $\chi \in R^{|S|}$ corresponds to the total cost of a scenario.

Using the cost vector χ, both SDORS-CVaR and SDORS-EV models are solved for the performance comparison between the models. Hence, the average ($\bar{\chi}$) and the variance (ν) of the cost vectors are calculated for each model. Unlike the EV, the CVaR allows the user to develop a schedule based on the preferred level of confidence for reducing variability of

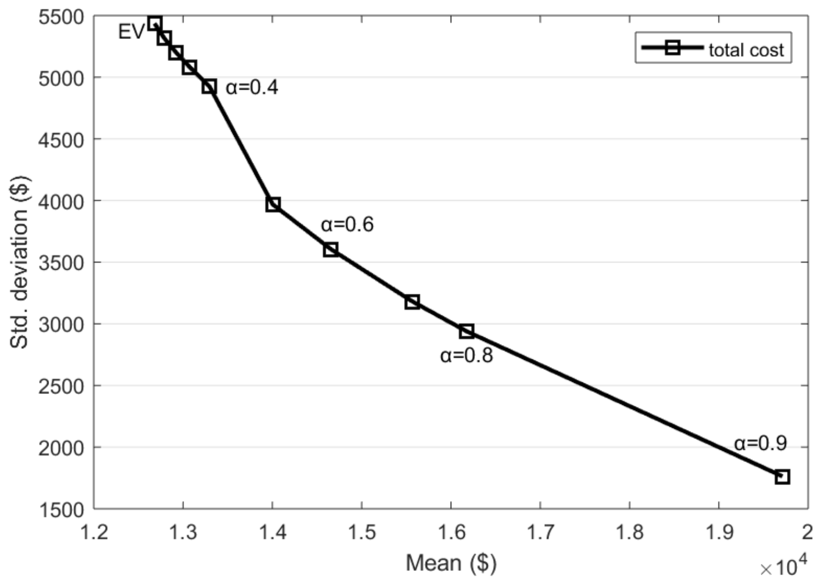


Figure 3.2: Pareto frontier schedules given by the SDORS-CVaR model

the random parameter. Figure 3.3 illustrates the mean-variance comparison between CVaR and EV, using four different probability levels, $60\%, 70\%, 80\%$, and 90%. One can use ratio $\bar{\chi}_{CVaR}/\bar{\chi}_{EV}$ to show the relative performance comparison between the CVaR-based model and the EV model. The ratio of 1 implies that both models perform about the same in minimizing the average total cost. If the ratio is less than 1, it can be interpreted that CVaR performs better than EV, and *vice versa*. Similarly, the ratio $\nu_{CVaR}/\nu_{EV} = 1$ implies as the optimal cost vectors of CVaR and EV have equal variances.

Our numerical results show that CVaR outperformed EV in reducing the variance of total cost. For each probability level, CVaR lowered the variance of the cost by 37% compared to EV on average. In 29 out of 32 cases (90.6%), the variance reduction resulted from the CVaR model was 20% or more than that of EV. Furthermore, CVaR produced schedules with at least 40% lower variance than EV in 37% of the cases (12 out of 32). More importantly, the CVaR significantly outperformed the EV in reducing the variance at a slight increase of the expected value of the total cost. We compared the performance of the CVaR and EV in minimizing the average total cost using the ratio $\left(\frac{\bar{\chi}_{CVaR}-\bar{\chi}_{EV}}{\bar{\chi}_{CVaR}}\right)$. The average value of the ratio from all 32 cases was 3.625%. Overall, CVaR demonstrated a superior performance by producing OR schedules with a lower variance of cost. On average, CVaR contributed a 37% lower variance at an increase of 3.6% in the average cost compared to EV. Note that CVaR can perform well in average cost savings if reducing the worst-case scenarios is the primary interest in OR scheduling.

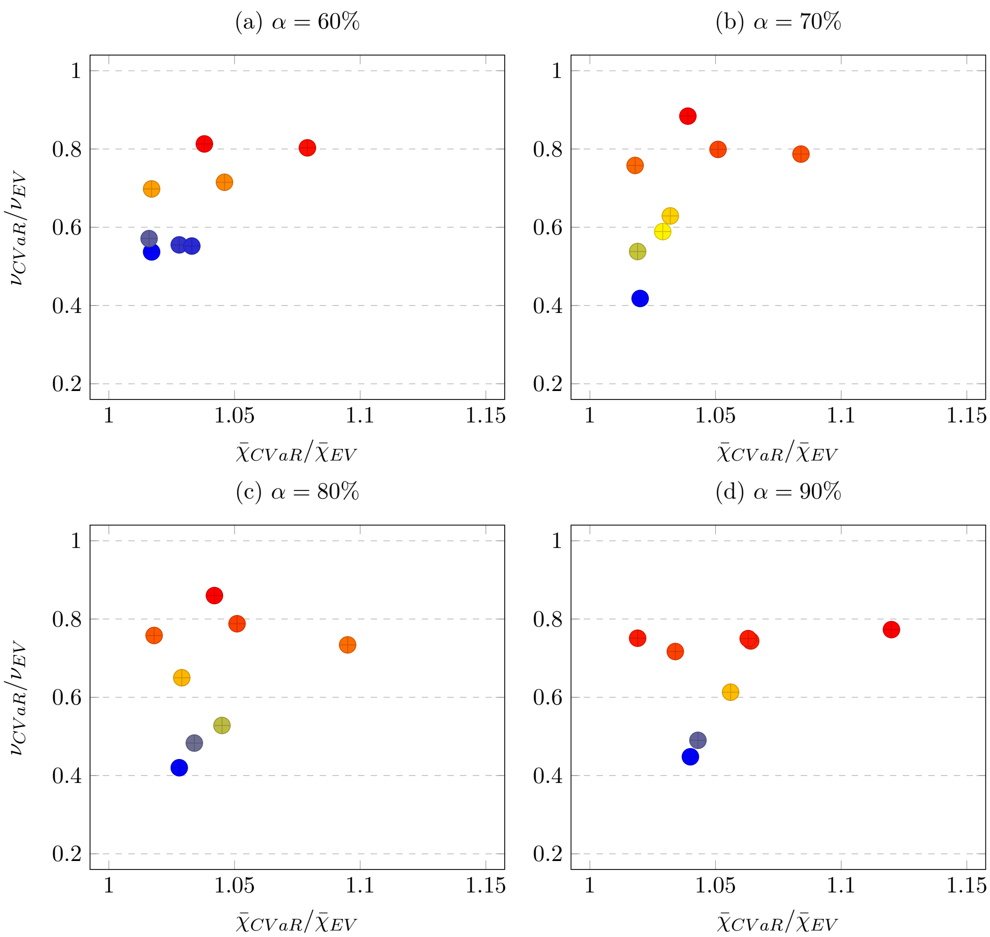


Figure 3.3: CVaR vs. EV: total cost mean-variance comparison

Risk-averse (i.e., conservative) decision-makers are often interested in minimizing the cost in the worst-case rather than the average value. The SDORS-CVaR model is designed to address such a case better than the EV model because the EV is a *risk-neutral* approach. Figure 3.4 shows the cost savings for the worst $(1-\alpha)\%$ of the scenarios when CVaR is selected over EV. The savings are calculated using the following formulas

$$Savings = \frac{E\left(\chi_{EV}|F(\chi_{EV}) > \alpha\right) - E\left(\chi_{CVaR}|F(\chi_{CVaR}) > \alpha\right)}{E\left(\chi_{EV}|F(\chi_{EV}) > \alpha\right)} \times 100\%$$

where $F(.)$ is the CDF of a cost vector. Figure 3.4 illustrates the savings for all instances using two different probability levels, 80% and 90%, to show how CVaR corresponds to a different confidence level. When $\alpha = 80\%$, CVaR results in cost savings between 1.47-4.25%. Also for $\alpha = 90\%$, cost savings range between 3.28-7.16%. The average savings percentage for $\alpha = 80\%$ and 90% are 2.52% and 4.04%, respectively. The cost savings tend to increase as the problem size grows. Figure 3.4 shows that if the decision-maker is focused on minimizing the worst outcomes, using the SDORS-CVaR model with higher probability levels is preferred. Moreover, the SDORS-CVaR model resulted in lower worst-case costs than the EV; CVaR reduces the worst-case cost by 9.36% on average.

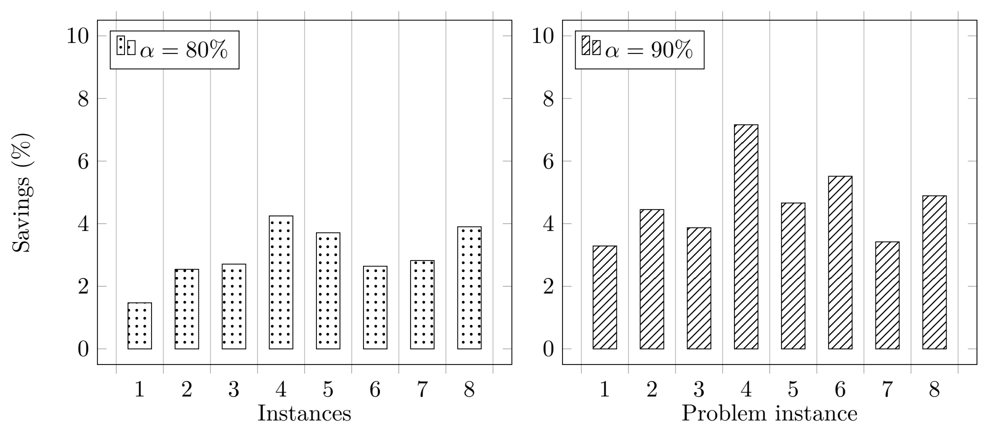


Figure 3.4: CVaR vs. EV: average cost savings in the worst $(1 - \alpha)\%$ of the scenarios

3.4.3 Discussions on Variability Reduction

Previously we showed that using CVaR produces superior results to EV in terms of reducing the variance of the total cost. However, variance and standard deviation are called non-robust measures of scale because outliers have a large impact on such measures. In this section, we use the widely used *robust* measures of scale, the interquartile range (IQR) and the median absolute deviation (MAD) to validate the ability of the SDORS-CVaR model in reducing the variability of cost. All instances are solved using the SDORS-CVaR model with four different probability levels, and the SDORS-EV model. For each instance and probability level, the IQR and MAD of the cost vectors obtained by CVaR are divided by that of the EV. Then, the ratios are plotted on Figure 3.5 and Figure 3.6. It is observed that for all instances and probability levels, CVaR achieves lower IQR and MAD values than EV for the cost vector. In more than half of the cases, the variability of the cost vector produced by CVaR is at least 20% lower than that of the EV. On average, CVaR results in cost vectors with 25.2% lower IQR and 24.4% lower MAD than EV. It is observed that setting the probability level to 60% and 70% in the SDORS-CVaR model leads to the largest variability reductions in the cost vector, where CVaR achieves 32.95% lower IQR and 33.65% lower MAD than EV on average.

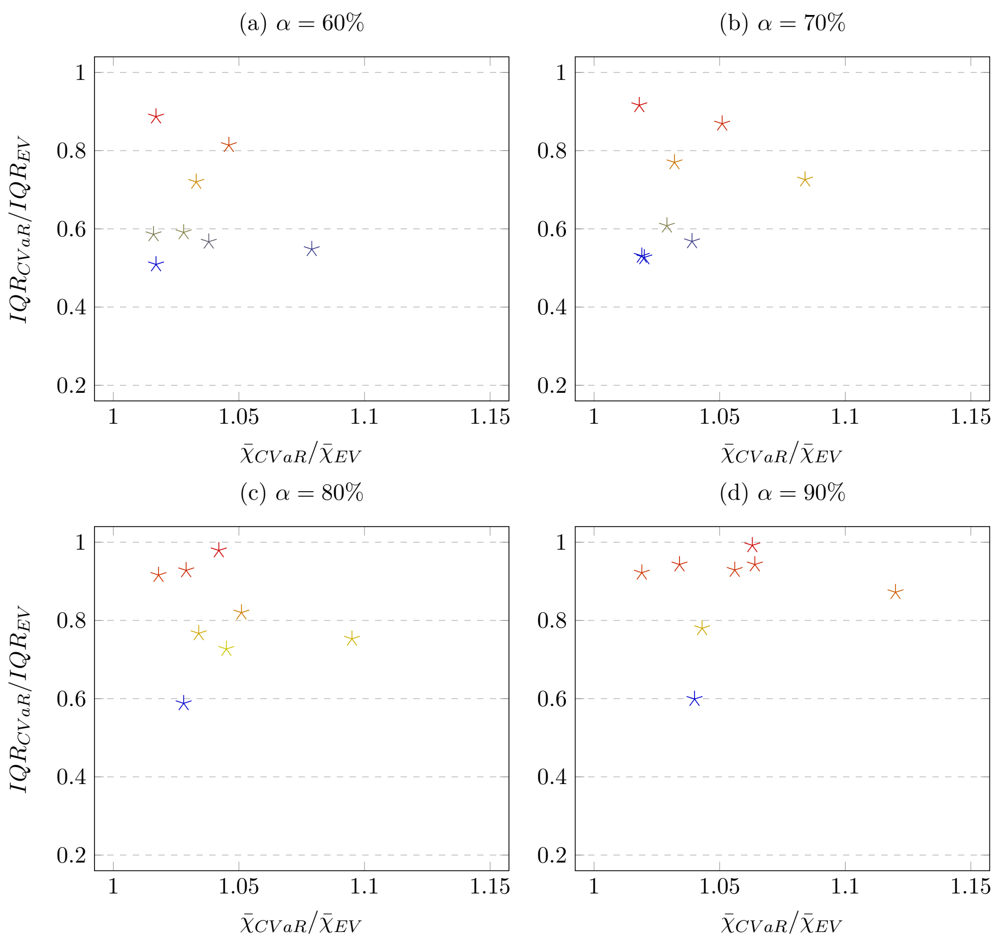


Figure 3.5: CVaR vs. EV: total cost mean-variance comparison using IQR

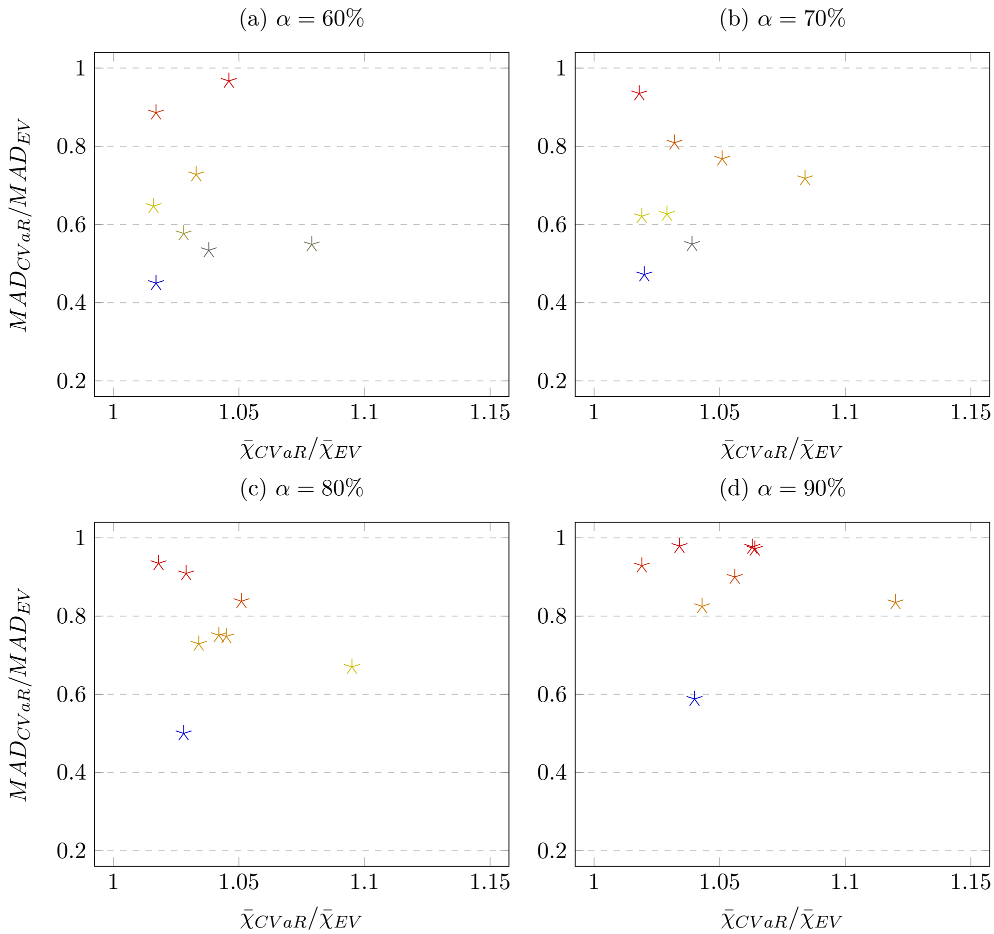


Figure 3.6: CVaR vs. EV: total cost mean-variance comparison using MAD

Figure 3.7 depicts the PDF plots of the cost vectors obtained by CVaR and EV for the instances {2,4,6,8}. The CVaR results are shown for two different probability levels. The PDF plots are estimated using a normal kernel function in MATLAB. The estimated PDF plots shown below add further proof that the SDORS-CVaR model shows promise for reducing the variability of the cost. It is also noticed that choosing higher probability levels in the SDORS-CVaR model results in larger reductions in the worst-case cost.

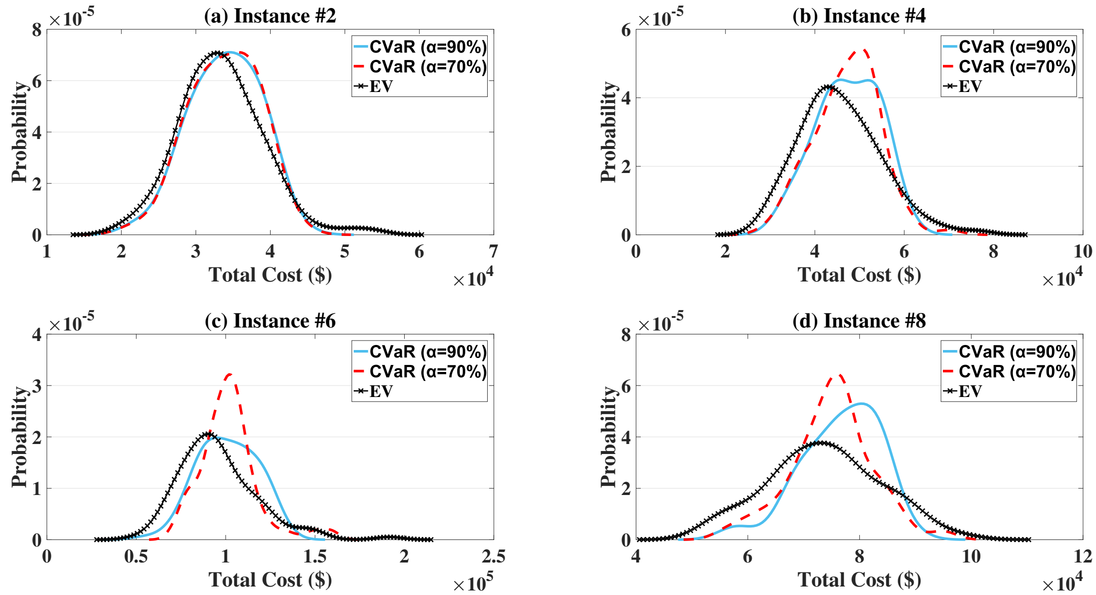


Figure 3.7: CVaR vs. EV: variability comparison using PDF estimates

3.4.4 Computational Complexity of SDORS-CVaR

It has been shown that the surgical case scheduling problem is NP-hard [11]. The optimization model is extremely difficult to solve due to the corresponding model having a large number of integer variables. For example, even the deterministic version of the SDORS-CVaR model contains more than 200,000 binary variables for a test problem instance solved in this chapter. Table 3.2 show the computational complexity of the SDORS-CVaR model when 100 scenarios are used for each problem instance.

Table 3.2: Computational complexity of SDORS-CVaR model

Problem Instance ID	1	2	3	4	5	6	7	8
Number of Patients	9	22	34	40	63	74	89	101
Number of ORs	5	5	5	10	20	20	15	20
Binary variables	405	2,420	5,950	16,400	80,640	111,000	120,150	206,040
Constraints	40,040	280,748	747,660	2,171,422	12,622,268	18,698,210	22,092,250	40,408,424

We also compared the computational efficiency of the SDORS-CVaR model with SDORS-EV in solving the problem instances. Table 3.3 shows the computation times for problem instances that are solved to optimality within a one-hour time limit. We added two more problem instances, containing 16 and 31 patients, that are numbered 9 and 10 in the table, respectively. It is observed that the the SDORS-CVaR model converges to the optimal solution faster than the SDORS-EV model as the problem size grows. For example in problem instance 3, the SDORS-CVaR with $\alpha = 90\%$ finds the optimal schedule within 8 minutes while the SDORS-EV fails to do so within one hour.

Table 3.3: CVaR vs. EV: solution time comparison

Problem Instance	CPU Time (sec)			Optimality Gap (%)		
	EV	CVaR ($\alpha = 50\%$)	CVaR ($\alpha = 90\%$)	EV	CVaR ($\alpha = 50\%$)	CVaR ($\alpha = 90\%$)
1	3.36	2.82	2.32	0.0	0.0	0.0
2	3352	1868	389	0.0	0.0	0.0
3	3600*	2909	451	4.5	0.0	0.0
9	2.5	4.2	2.3	0.0	0.0	0.0
10	1065	119	99	0.0	0.0	0.0

* Stopped due to exceeding the solution time limit and not solved to optimality.

3.5 Conclusion

A risk-based solution method is proposed for the daily OR scheduling problem with stochastic surgery durations. The problem was modeled using a scenario-based MILP formulation to reduce computational burden to solve the model. Instead of using the commonly used expected value-based method, we have developed a CVaR-based approach to control variability on overtime and idle time, and reduce worst-case outcomes of an OR schedule in terms of cost. The proposed solution method was tested using a set of real-life numerical instances. In all test cases, the CVaR outperformed the EV in reducing variability on overtime, idle time, and the associated worst-case costs. As compared to the EV, the CVaR reduced variance on total cost by 37%, produced a 25% lower interquartile range and 24% lower median absolute deviation while increasing the average cost by less than 4%. Furthermore, we have introduced an adjustable confidence level parameter that allows a decision-maker to be able to switch the emphasis between reducing variability and reducing the average of total cost associated with overtime and idle time. When the region of interest is narrowed by selecting higher confidence levels (e.g., 90%), it resulted in better performance in terms of reducing the variability.

Chapter 4

A Decomposition Algorithm for the Two-Stage Chance-Constrained Operating Room Scheduling Problem

4.1 Introduction

Health care expenditures are expected to constitute 25% of the US gross domestic product (GDP) in 2025, an increase from 15.9% in 2005 [66]. Surgical expenses contribute to 30% of health care expenditures and are expected to grow from $572 billion in 2005 to $912 billion (2005-valuated dollars) in the year 2025. Surgical procedures are complex tasks requiring a variety of specialized and expensive resources. In 2011, hospitalizations involving surgical procedures constituted 29% of total hospital stays while contributing to 48% of total hospital costs in the US [94]. In light of these reports, surgeries are recognized as the most crucial activities performed in hospitals from a social, medical and economic point of view.

Several survey articles have recognized the surgery duration uncertainty as a major obstacle to developing practical and cost-effective OR schedules [61]. This chapter proposes a chance-constrained programming model that: 1) provides cost-effective OR schedules by considering both deterministic and stochastic costs, 2) maintains a low OR overtime probability and compares individual and joint chance constraints, 3) results in a better cost-variability trade-off compared to two existing models in the literature, and 4) solves such problems at a faster rate than the two aforementioned existing models before applying any solution algorithms. Moreover, a computationally efficient solution method and strong valid inequalities are provided to facilitate timely decision-making in the case of disruptions in the schedule.

The OR scheduling literature has been reviewed in several survey articles [13, 61, 79]. The published literature has been classified using several categories, including *uncertainty*. Variable surgery duration is one of the most commonly studied sources of uncertainty by the Operations Research community. It is shown that mitigating the impact of disruptions in the schedule due to uncertainty can lead to higher capacity utilization and lower costs [62, 65]. Hence, it is crucial to ensure that the provided schedule works reliably in the presence of large variability in surgery durations.

Numerous works have used stochastic programming to model the uncertain surgery durations in the OR scheduling problems [5, 29, 73, 93]. The majority of these models consider optimizing the expectation of costs/revenues during the planning horizon [36]. For problems with moderate variability, using the expected value (EV) can result in desirable outputs. However, the obtained solutions may show poor performance for problems displaying frequent changes in a less predictable manner [72]. A number of articles considered using the Conditional Value-at-Risk (CVaR) [78] to account for undesirable realizations of the uncertain parameters [52, 67, 80]. The CVaR function minimizes the expected tail of costs.

Another array of articles have used the chance-constrained programming (CCP) models [15] to address the uncertainty. This approach mitigates the risk of disadvantageous events (e.g., OR overtime, patient waiting time) exceeding the specified thresholds, rather than merely minimizing their expected value [19, 42, 97]. Shylo, Prokopyev and Schaefer [84] applied chance-constraints to control the OR block overtime in the OR surgery planning problem. Zhang, Denton and Xie [98] studied a chance-constrained OR surgery allocation problem. Deng and Shen [19] developed a two-stage stochastic model for the multi-server appointment scheduling problem with a joint chance constraint on server overtime. They applied the proposed model and solution approach to solve OR scheduling problem test instances. Jebali and Diabat [40] studied the surgery planning problem under uncertain surgery duration, length of stay in the intensive care unit (ICU), and emergency patient arrival. They employed chance constraints to control the violation of ICU capacity. Wang, Li and Peng [92] proposed a distributionally robust chance-constrained model for the surgery planning problem with stochastic surgery durations. Noorizadegan and Seifi [71] proposed a CCP model for the surgery planning problem with uncertain surgery durations. Kamran, Karimi and Dellaert [41] proposed a two-stage stochastic model with chance constraints on OR overtime for the advance scheduling problem. Deng, Shen and Denton [21] developed a distributionally robust chance-constrained model for the OR scheduling problem. They control the risk of OR overtime and surgery waiting using joint chance constraints.

Table 4.1 summarizes the selected published literature and identifies the research gaps that are addressed in this chapter. First, it can be observed that very few articles proposed a CCP model for the OR scheduling problem under uncertainty [19, 21]. Surgery scheduling problems often have a more complex structure resulting from a variety of decisions, such as OR opening, patient-to-OR assignment, surgery sequencing, and projected and actual start times before and after the realization of random surgery durations, respectively. CCP-based models have the potential to effectively handle such large variabilities in daily surgery scheduling problems [61].

Table 4.1: Research Gaps in the Chance-Constrained OR Scheduling Literature

Authors (Year)	**Decisions**			**Objective**	
	OR	**Resource Allocation**	**Sequencing & Scheduling**	**Deterministic Costs**	**Stochastic Costs**
Shylo et al. (2012)		✓			OR Idle Time
Zhang et al. (2015)		✓			OR Overtime
Deng et al. (2019)	✓	✓	✓	OR Opening	
Noorizadegan & Seifi (2018)	✓	✓	✓	OR Opening & Turn-Over	
Wang et al. (2017)	✓	✓		Operational	
This Chapter	✓	✓	✓	OR Opening	OR Overtime & Patient Waiting Time

Second, a majority of the models have neglected the importance of minimizing the stochastic second-stage costs. Their primary focus has been on providing schedules within the specified risk tolerances while also aiming to minimize deterministic performance measures, such as fixed OR opening costs [21, 71]. Unlike existing approaches, this chapter proposes a chance-constrained model that aims to minimize both deterministic and stochastic costs for the OR scheduling problem. The significance of considering both classes of costs is highlighted using numerical experiments.

Third, we provide insightful observations about the performance of three different models (CCP, CVaR and EV) in solving the stochastic OR scheduling problem under various risk thresholds. The proposed model is compared alongside EV and CVaR models using several metrics such as total costs, OR utilization and solution time. Moreover, the performances of both individual and joint chance constraints are compared in terms of OR opening decisions, minimizing the second-stage costs and computational efficiency.

Finally a computationally efficient decomposition algorithm is applied to provide high-quality solutions for the large-scale test instances within reasonable time frames. We proposed an algorithm to derive feasibility cuts using the first-stage solutions that accelerate finding feasible solutions and the convergence speed.

4.2 Mathematical Formulation

4.2.1 Problem Description

Let I be the set of elective surgeries and R be the set of operating rooms. The problem is to schedule surgeries over a daily planning horizon. We assume that operating rooms use the block booking policy [25]. Each OR is allocated to a surgical specialty according to the master surgery schedule. The incidence matrix $E = \{e_{ir}\}$ for all $i \in I, r \in R$ allows specific surgery-to-OR assignments. The surgery duration is random, denoted by a random vector $\delta = (\delta_1, ..., \delta_{|I|})^T \in \mathbb{R}_+^{|I|}$ where δ_i shows the random duration for surgery $i \in I$. We assume that the random surgery duration has finite and discrete support S for δ. The probability density of each scenario s is denoted by p_s, where $\sum_{s \in S} p_s = 1$. Each realization of δ in scenario s is shown by $\delta^s = (\delta_1^s, ..., \delta_{|I|}^s)^T$. Every surgery on the daily booking list must be operated. No interruption is allowed once an operation has started. It is desired to decrease the probability of working OR overtime. This restriction is enforced using chance constraints. Each surgery must be assigned an OR and a projected start time. Our goal is to minimize the sum of fixed OR opening costs and expected costs corresponding to OR overtime and patient waiting times.

4.2.2 Chance-Constrained OR Scheduling Problem

We propose a two-stage stochastic model for the daily OR scheduling problem with individual chance constraints enforcing overtime restrictions on every OR. Table 4.2 introduces the notation used in our model.

Table 4.2: Sets, parameters and variables used in the model

	Symbol	**Definition**
Indices	i	index for elective surgeries, $i \in I$
	k	index for order of surgery appointments in OR, $k \in K$
	r	index for operating rooms, $r \in R$
	s	index for scenarios, $s \in S$
Parameters	e_{ir}	1, if surgery i can be assigned to OR r; 0, otherwise
	δ_{is}	duration of surgery i in scenario s
	f_r	fixed cost of opening OR r
	cap_r	operating time limit for OR r
	co_r	unit overtime cost of OR r
	cw_r	unit waiting cost for surgery i
	p_s	probability density of scenario s
	α_r	overtime probability threshold for OR r (confidence level)
	M	a sufficiently large number
Decision Variables	u_r	1, if OR r is opened; 0, otherwise
	y_{ikr}	1, if surgery i scheduled as k^{th} surgery in OR r; 0, otherwise
	z_{rs}	1, if chance constraint on OR r is violated in scenario s; 0, otherwise
	tp_{kr}	projected start time for surgery i
	tr_{krs}	actual start time for surgery i in scenario s
	o_{rs}	OR r overtime in scenario s
	w_{rs}	total patient waiting times in OR r and scenario s

The first-stage problem involves deterministic decision-making (i.e., OR opening and surgery case assignment) prior to the realization of uncertain surgery durations. After the uncertain parameters are revealed, the second-stage problem determines recourse actions (e.g., adjusting start times and adding OR overtime) that incur additional costs to provide meaningful schedules based on first-stage decisions. The goal is to minimize total costs as well as satisfy the chance constraints on OR overtime.

The first-stage problem (M_1) can be formulated as

$$Min \sum_{r \in R} f_r u_r \tag{4.2.1}$$

$$\text{and} \quad x \in X \tag{4.2.2}$$

where $x = (u, y)$ is the vector of first-stage variables. Set X is the resulting set from the deterministic constraints (4.2.3)-(4.2.7) formulated as

$$y_{ikr} \leq u_r \qquad \forall i \in I, k \in K, r \in R, \tag{4.2.3}$$

$$\sum_{k,r} y_{ikr} = 1 \qquad \forall i \in I, \tag{4.2.4}$$

$$y_{ikr} \leq e_{ir} \qquad \forall i \in I, k \in K, r \in R, \tag{4.2.5}$$

$$\sum_i y_{i(k+1)r} \leq \sum_i y_{ikr} \qquad \forall k \in K \setminus \{|K|\}, r \in R, \tag{4.2.6}$$

$$\text{and} \quad u_r \in \mathbb{B}^{|R|}, y_{ikr} \in \mathbb{B}^{|I| \times |K| \times |R|} \qquad \forall i \in I, k \in K, r \in R. \tag{4.2.7}$$

Objective function (4.2.1) minimizes the total cost of opening operating rooms. Constraints (4.2.3) and (4.2.4) ensure that every surgery will be assigned to one and only one spot in an open OR during the day. Constraint (4.2.5) enforces eligible surgery-to-OR assignments. Constraint (4.2.6) determines the order of operating surgical cases in each OR. Constraint (4.2.7) enforces binary values for the first-stage decision variables.

The chance-constrained second-stage problem (M_2) is formulated as

$$tp_{kr} \leq tp_{(k+1)r} \qquad \forall k \in K \setminus \{|K|\}, r \in R, \tag{4.2.8}$$

$$tp_{kr} \leq tr_{krs} \qquad \forall k \in K, r \in R, s \in S, \tag{4.2.9}$$

$$tr_{krs} \leq tr_{(k+1)rs} \qquad \forall k \in K, r \in R, s \in S, \tag{4.2.10}$$

$$\sum_k (tr_{krs} - tp_{kr}) \leq w_{rs} \qquad \forall r \in R, s \in S, \tag{4.2.11}$$

$$Pr\{tr_{krs} + \sum_i \delta_{is} y_{ikr} \geq cap_r \qquad \forall k \in K, s \in S\} \leq \alpha_r, \forall r \in R, \tag{4.2.12}$$

$$\text{and} \quad tp_{kr}, tr_{krs}, w_{rs} \geq 0 \qquad \forall k \in K, r \in R, s \in S. \tag{4.2.13}$$

Constraint (4.2.8) determines the projected start time for each surgery according to the sequencing decisions. Constraint (4.2.9) ensures that each surgery starts after its projected start time. Constraint (4.2.10) is similar to (4.2.8) in that the actual start times must follow the sequencing decisions. Constraint (4.2.11) calculates the amount of waiting time in every OR per scenario. The chance constraints (4.2.12) state that the surgeries assigned to an OR must be finished during the regular hours (i.e., no overtime) with high probability. Constraint (4.2.13) enforces the non-negativity of the second-stage decision variables. The objective function of the second-stage problem is formulated in the remainder of this section.

The set $\mathbb{P}(s)$ of the first-stage solutions that are made to satisfy the chance-constrained second-stage problem is derived as

$$\mathbb{P}(r, s) = \left\{ x \in X \mid \exists tp, tr : tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r \right\} \tag{4.2.14}$$

$$\text{and} \quad \mathbb{P}(s) = \bigcap_{r \in R} \mathbb{P}(r, s). \tag{4.2.15}$$

Proposition 1. *Let $\alpha_r |S|$ be an integer for every r. Then, chance constraints* (4.2.12) *are equivalent to*

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + M z_{rs} \qquad \forall k \in K, r \in R, s \in S \tag{4.2.16}$$

$$\textit{and} \quad \sum_s z_{rs} \leq \alpha_r |S| \qquad \forall r \in R, s \in S. \tag{4.2.17}$$

where binary variable $z_{rs} = 1$ when the time capacity of room r is violated.

Proof. According to constraints (4.2.12) and (4.2.13), an overtime occurs if

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} > cap_r, \forall k \in K, r \in R, s \in S. \tag{4.2.18}$$

Therefore, the value of z_{rs} in (4.2.16) captures scenarios where an OR runs overtime. Given that the random surgery duration has a discrete and finite support, constraint (4.2.17) limits the number of scenarios where each OR can run overtime. $\square$

For each scenario $s \in S$, an operation may be completed during regular hours ($z_{rs} = 0$) or may run into overtime ($z_{rs} = 1$). Therefore, the second-stage cost, $g(x, s)$, will be calculated differently in each case

$$\begin{cases} g_r^1(x,s) = cw_r w_{rs} & z_{rs} = 0 \quad (4.2.19) \\ g_r^2(x,s) = cw_r w_{rs} + co_r o_{rs} & z_{rs} = 1 \quad (4.2.20) \end{cases}$$

where o_{rs} is a non-negative variable representing overtime. Therefore, the objective function of the second-stage problem can be formulated as

$$\mathbb{E}_s\left[g(x,s)\right] = \mathbb{E}_s\left(\sum_r (1 - z_{rs}) g_r^1(x,s) + z_{rs} g_r^2(x,s)\right). \tag{4.2.21}$$

Objective function (4.2.21) minimizes the expected costs corresponding to OR overtime and patient waiting times. The deterministic equivalent formulation for the two-stage chance-constrained OR scheduling model (M_{DEF}) can be modeled as

$$Min \quad obj = \sum_r f_r u_r + \frac{1}{|S|}\sum_{r,s}\left(cw_r w_{rs} + co_r o_{rs}\right), \tag{4.2.22}$$

$$s.t. \quad (4.2.8) - (4.2.11), (4.2.13),$$

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + M z_{rs} \qquad \forall k \in K, r \in R, s \in S, \tag{4.2.23}$$

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + M(1 - z_{rs}) + o_{rs} \qquad \forall k \in K, r \in R, s \in S, \tag{4.2.24}$$

$$\sum_s z_r^s \leq \alpha_r |S| \qquad \forall r \in R, s \in S, \tag{4.2.25}$$

$$\text{and} \quad x \in X, z_{rs} \in \mathbb{B}^{|R|\times|S|} \qquad \forall k \in K, r \in R, s \in S. \tag{4.2.26}$$

Using a big M value can lead to weak LP relaxations. Assigning a smaller value for M can help tighten the feasible region for the LP relaxation of M_{DEF}. Therefore, instead of setting a single large value for M, constraint-specific formulae used to calculate the big M values are developed for constraints (4.2.23) and (4.2.24) as

$$M_{rs} = \sum_i e_{ir}\delta_{is}, \forall r \in R, s \in S. \tag{4.2.27}$$

The values in (4.2.27) are valid because the total operation time in each OR does not exceed the duration of all surgical cases that can be allocated to the specific operating room.

4.3 Solution Approach

This section describes a decomposition algorithm that solves the proposed model in Section 4.2. The proposed algorithm can solve the model to optimality if the following assumptions are satisfied [53, 54]:

- The random vector S has discrete and finite support. Specifically, $p_s = \frac{1}{|S|}$ for $s \in S$.

  We have stated this assumption in the problem description in Section 4.2.1.

- Set X and $\mathbb{P}(s)$, $s \in S$ are non-empty compact sets.

  Without loss of generality, we can assume that for every $s \in S$, there exists a feasible first-stage solution that satisfies the chance constraints. Therefore, sets X and $\mathbb{P}(s)$ are finite sets of points that qualifies them as compact sets.

- Set $conv\left(\mathbb{P}(s)\right)$, $s \in S$ have the same recession cone, i.e., there exists $C \subseteq \mathbb{R}^N$ such that $C = \left\{\theta \in \mathbb{R}^N | x + \lambda\theta \in \mathbb{P}(s); \forall x \in \mathbb{P}(s), \lambda \geq 0\right\}$ for all $s \in S$, where $N := |R| + |I| \times |K| \times |R|$.

  Proof. Since all of the first-stage decision variables are binary, $\mathbb{P}(s)$ is bounded by a hypercube of dimension N. Therefore, $\theta = 0$ is the only solution that satisfies the condition in the definition of C. In other words, $C = \{0\}$ for all $s \in S$. □

- There does not exist an extreme ray $\tilde{\theta}$ of conv(X) with $f^T\tilde{\theta} < 0$, i.e., the two-stage problem has a bounded optimal solution.

Proof. We need to show that both first-stage and second-stage problems have bounded optimal solutions. We know from Assumption 2 that both problems are feasible. The highest objective function value for the first-stage problem is when all operating rooms are open, i.e., $\sum_r f_r$, which is bounded. Given the first-stage solution, every open operating room will run overtime in up to $\alpha_r|S|$ scenarios. The amount of overtime is bounded by $\max_{k\in K}\{tr_{krs} + \sum_i \delta_{is}y_{ikr} - cap_r\}$, which is also a finite value given $tr_{krs} \geq 0$ and the minimization objective function in the second-stage problem. □

Proposition 2. *Model M_{DEF} has an optimal solution (x^*, z^*) in which $\sum_s z^*_{rs} = \alpha_r|S|$ for all $s \in S$ [54].*

Proof. This holds for the individual chance constraints in our model without loss of generality. Assume that in the optimal solution to our model, there exists $r \in R$ where $\sum_s z^*_{rs} = \epsilon < \alpha_r|S|$. The optimal solution will allow $(\alpha_r|S| - \epsilon)$ scenarios to run in overtime mode ($z_{rs} = 1$) with the corresponding overtime variables o_{rs} set to zero. □

We begin the decomposition algorithm by defining feasibility ($\mathbb{F}$) and optimality ($\mathbb{O}$) sets as

$$\mathbb{F} = \left\{ x \in X, z \in \mathbb{B}^{|R|\times|S|} : \sum_s z_{rs} = \alpha_r|S|, r \in R, z_{rs} = 0 \Rightarrow x \in \mathbb{P}(r,s), s \in S \right\}, \quad (4.3.1)$$

$$\mathbb{O} = \left\{ (x, z, \rho) \in \mathbb{F} \times \mathbb{R}_+ : \rho \geq \frac{1}{|S|}\sum_{r,s}(1 - z_{rs})g^1_r(x,s) + z_{rs}g^2_r(x,s) \right\}. \quad (4.3.2)$$

These sets will be approximated using feasibility and optimality cuts in the following master problem (**MP**)

$$\begin{aligned} &\underset{x,z,\rho}{Min} \quad f^T u + \rho, && (4.3.3)\\ &\sum_s z_{rs} = \alpha_r|S| && \forall r \in R,\\ &x \in X, z \in \mathbb{B}^{|R|\times|S|}, \rho \geq 0, \end{aligned}$$

$$(x, z) \in \tilde{\mathbb{F}},$$

$$\text{and} \quad (x, z, \rho) \in \tilde{\mathbb{O}}.$$

The sets $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{O}}$ are the outer approximations of the feasibility (F) and optimality (O) sets, respectively. In the remainder of this section, we will derive strong valid inequalities to define $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{O}}$.

4.3.1 Feasibility Cuts

Two sets of subproblems are required to formulate the strong feasibility cuts: single-scenario optimization and single-scenario separation [53]. The optimization subproblem for the OR scheduling problem is formulated as

$$h_{rs}(\gamma) = \underset{x}{Min} \left\{ \gamma x \mid x \in \mathbb{P}(r,s) \cap \bar{X} \right\} \tag{4.3.4}$$

where $\gamma \in \mathbb{R}^N$ and $\bar{X} \supseteq X$, such that $\mathbb{P}(r,s) \cap \bar{X} \neq \Phi$.

Proposition 3. *Problem* (4.3.4) *is feasible and has a finite optimal value if* $\gamma \in \mathbb{R}^N$.

The separation subproblem can be formulated as

$$\varrho_{rs}(\hat{x}) = \underset{\pi}{Max} \sum_k \pi^4_{krs} \left(\sum_i \tilde{d}_{is} \hat{y}_{ikr} - cap_r \right), \tag{4.3.5}$$

$$\left(\pi^1_{(k-1)rs} - \pi^1_{krs} \right) + \pi^5_{rs} \leq 0 \qquad \forall k \in K, \tag{4.3.6}$$

$$\left(\pi^2_{(k-1)rs} - \pi^2_{krs} \right) + \pi^3_{krs} - \pi^4_{krs} - \pi^5_{rs} \leq 0 \qquad \forall k \in K, \tag{4.3.7}$$

$$\pi^4_{krs} \leq co_r \qquad \forall k \in K, \tag{4.3.8}$$

$$\pi^5_{rs} \leq cw_r, \tag{4.3.9}$$

$$\sum_k \left(\pi^1_{krs} + \pi^2_{krs} + \pi^3_{krs} + \pi^4_{krs} \right) + \pi^5_{rs} = 1, \tag{4.3.10}$$

$$\text{and} \quad \pi^1_{krs}, \pi^2_{krs}, \pi^3_{krs}, \pi^4_{krs}, \pi^5_{rs} \geq 0 \qquad \forall k \in K. \tag{4.3.11}$$

Solving this subproblem to optimality returns a separating hyperplane of the form $\gamma x \geq$

β for all $x \in \mathbb{P}(r,s)$. Let $\hat{x}$ be a solution to the master problem (MP). If $\varrho_{rs}(\hat{x}) > 0$ and $\hat{\pi}$ is the optimal solution, the separating hyperplane $-\sum_k \hat{\pi}_k^4 \sum_i \tilde{d}_i y_{ikr} \geq -\sum_k \hat{\pi}_k^4 cap_r$ cuts off $\hat{x}$ from $\tilde{\mathbb{F}}$. Therefore, we define the valid feasibility cuts as

Theorem 4.3.1. *The following sets of inequalities are valid for* $\mathbb{F}$*:*

$$\gamma x + \sum_{i=1}^{l} \left(h_{g_i}(\gamma) - h_{g_{i+1}}(\gamma)\right) z_{g_i} \geq h_{g_1}(\gamma) \tag{4.3.12}$$

where $h_{\sigma_1} \geq h_{\sigma_2} \geq ... \geq h_{\sigma_{|\mathbb{S}|}}$, $G = \{g_1, g_2, ..., g_l\} \subseteq \{\sigma_1, \sigma_2, ..., \sigma_p\}$ *and* $h_{g_{l+1}} = h_{\sigma_{p+1}}$.

The remainder of this section presents another class of feasibility cuts derived from the solutions to the first-stage problem (M_1) or master problem (MP) that result in the violation of chance constraints (4.2.12).

Theorem 4.3.2. *The following set of inequalities are valid for* $\mathbb{F}$

$$z_{rs} - \sum_{i \in TB, k} y_{ikr} \geq 1 - |TB| \quad \forall r \in R, s \in S \tag{4.3.13}$$

where TB *is the subset of surgeries that lead to the violation of chance constraints when they are assigned to the same OR.*

Let $\hat{y}_{ikr}$ be the set of surgery-to-OR assignments obtained from solving (M_1) or (MP). Algorithm 1 generates inequalities of type (4.3.13).

Algorithm 1 Feasibility Cut Generation

input: Sets and parameters in Table 4.2, $\hat{y}_{ikr}$.

initialize: $v_{rs} \leftarrow 0$, $Cnt_r \leftarrow 0$, $TB \leftarrow \Phi$.

Calculate $v_{rs} = \sum_{k,i} \delta_{is}\hat{y}_{ikr} - cap_r, \forall r \in R, s \in S$.

if $v_{rs} > 0, \forall r \in R, s \in S$ **then**

  • $Cnt_r \leftarrow Cnt_r + 1$.

  if $Cnt_r > \alpha_r|S|, \forall r \in R$ **then**

    • $TB \leftarrow \{i|\hat{y}_{ikr} = 1\}$.

    • Add feasibility cut (4.3.13) to $\tilde{\mathbb{F}}$ in (**MP**).

  end if

end if

output: Feasibility cuts to add to the model (**MP**).

4.3.2 Optimality Cuts

In this section, we derive optimality cuts to add to $\tilde{\mathbb{O}}$. First, we formulate the dual problems for regular and overtime modes of the second-stage problem. For every $r \in R$ and $s \in S$ where $z_{rs} = 0$, the regular mode dual problem is formulated as

$$\nu^1_{rs}(\hat{x}) = \underset{\pi}{Max} \sum_k \pi^4_{krs} \left(\sum_i \tilde{d}_{is}\hat{y}_{ikr} - cap_r \right), \tag{4.3.14}$$

$$\left(\pi^1_{(k-1)rs} - \pi^1_{krs}\right) + \pi^5_{rs} \leq 0 \qquad \forall k \in K, \tag{4.3.15}$$

$$\left(\pi^2_{(k-1)rs} - \pi^2_{krs}\right) + \pi^3_{krs} - \pi^4_{krs} - \pi^5_{rs} \leq 0 \qquad \forall k \in K, \tag{4.3.16}$$

$$\pi^5_{rs} \leq cw_r, \tag{4.3.17}$$

$$\sum_k \left(\pi^1_{krs} + \pi^2_{krs} + \pi^3_{krs} + \pi^4_{krs}\right) + \pi^5_{rs} = 1, \tag{4.3.18}$$

$$\text{and} \quad \pi^1_{krs}, \pi^2_{krs}, \pi^3_{krs}, \pi^4_{krs}, \pi^5_{rs} \geq 0 \qquad \forall k \in K. \tag{4.3.19}$$

Since the only difference between the regular and overtime mode is the introduction of overtime variables (when $z_{rs} = 1$), the overtime mode dual problem can be derived by replacing π with $\bar{\pi}$ and adding the dual constraints corresponding to overtime variables as

$$\nu^2_{rs}(\hat{x}) = \underset{\bar{\pi}}{Max} \quad \sum_k \bar{\pi}^4_{krs}\left(\sum_i \tilde{d}_{is}\hat{y}_{ikr} - cap_r\right), \tag{4.3.20}$$

$$s.t. \quad (4.3.15) - (4.3.19),$$

$$\text{and} \quad \bar{\pi}^4_{krs} \leq co_r \qquad \forall k \in K. \tag{4.3.21}$$

The set of dual optimal solutions for the regular and overtime modes are shown by Π_{rs} and $\bar{\Pi}_{rs}$, respectively. Then, we formulate optimality subproblems for each mode. For a given $\tau \in \mathbb{R}^N$, $r \in R$ and $s \in S$, we formulate the optimality subproblem for the regular mode as

$$\psi^1_{rs}(\tau) = Min\left\{g^1_r(x,s) + \tau^T x : x \in \mathbb{P}(r,s)\right\}. \tag{4.3.22}$$

Similarly, we formulate the optimality subproblem for the overtime mode as

$$\psi^2_{rs}(\tau) = Min\left\{g^2_r(x,s) + \tau^T x : x \in X\right\}. \tag{4.3.23}$$

Proposition 4. *Let dom* $\psi_{rs}(\tau) = \left\{\tau \in \mathbb{R}^N : \psi_{rs}(\tau) > -\infty\right\}$. *There exists* $D \subseteq \mathbb{R}^N$ *where dom* $\psi^1_{rs}(\tau) =$ *dom* $\psi^2_{rs}(\tau) = D$.

Proof. From Assumption 4, we know that both (4.2.19) and (4.2.20) are non-negative and bounded. We also know that x is binary. It suffices to have $\tau \in \mathbb{R}^N_+$ such that $\psi^1_{rs}(\tau) > -\infty$ and $\psi^2_{rs}(\tau) > -\infty$. Therefore, $D = \mathbb{R}^N_+$ satisfies the condition. □

Proposition 5. *Let* $Q \subseteq S$, $\pi_{rs} \in \Pi_{rs}$ *and* $\tau_{rs} = \sum_{i,k}\pi^4_{krs}\tilde{d}_{is}$ *for* $s \in Q$, *and* $\bar{\pi}_{rs} \in \bar{\Pi}_{rs}$ *and* $\tau_{rs} = \sum_{i,k}\bar{\pi}^4_{krs}\tilde{d}_{is}$ *for* $s \in S \setminus Q$. *The following inequality is valid for* $\mathbb{O}$

$$\rho + \frac{1}{|S|}\sum_{r,s \in Q}\left(-\sum_k \pi^4_{krs}cap_r - \psi^2_{rs}(\tau_{rs})\right) z_{rs} \tag{4.3.24}$$

$$
\begin{aligned}
&+\frac{1}{|S|}\sum_{r,s\in S\setminus Q}\left(-\sum_{k}\bar{\pi}^{4}_{krs}cap_{r}-\psi^{1}_{rs}(\tau_{rs})\right)(1-z_{rs})\\
&\geq\frac{1}{|S|}\sum_{r,s\in Q}\left(-\sum_{k}\pi^{4}_{krs}cap_{r}\right)\\
&+\frac{1}{|S|}\sum_{r,s\in S\setminus Q}\left(-\sum_{k}\bar{\pi}^{4}_{krs}cap_{r}\right)+\frac{1}{|S|}\sum_{i,r,s}\left(\tau_{rs}\hat{y}_{ikr}\right).
\end{aligned}
$$

4.3.3 Decomposition Algorithm

A decomposition algorithm is proposed to solve the two-stage chance-constrained OR scheduling problem. This algorithm has a similar structure to the Benders decomposition algorithm [76]. Rather than traditional Benders cuts, we use strong valid inequalities derived in Section 4.3.1 and Section 4.3.2. Parameter ϵ in Algorithm 2 represents the upper bound on the relative optimality gap, calculated as $\frac{UB-LB}{UB}$.

Algorithm 2 Decomposition Algorithm

---

input: sets and parameters in Table 4.2, model M_{DEF}.
initialize: $LB := -\infty$, $UB := +\infty$, $\epsilon \in \left[10^{-3}, 10^{-6}\right]$
while $\frac{UB-LB}{LB} > \epsilon$ **do**
    Solve master problem (4.3.3).
    if (4.3.3) is infeasible **then**
        Stop. Original problem is infeasible.
    else
        Let $(\hat{x}, \hat{z}, \hat{\rho})$ be an optimal solution to (4.3.3).
        • $LB \leftarrow f^T\hat{u} + \hat{\rho}$.
        • Check feasibility of the second-stage problem by calling Algorithm 1 and evaluating the inequalities (4.3.12).
        if there exists violated inequalities **then**
            • Add feasibility cuts (4.3.12) and (4.3.13) to $\tilde{\mathbb{F}}$.
        else
            • $UB \leftarrow \sum_r f_r\hat{u}_r + \frac{1}{|S|}\sum_{r,s}\left[g^1_r(\hat{x},s)+g^2_r(\hat{x},s)\right]$
            • Add optimality cuts (4.3.24) to $\tilde{\mathbb{O}}$.
        end if
    end if
end while
output: optimal cost obj^* and decision variables $(x^*, z^*, tp^*, tr^*, o^*, w^*)$.

Theorem 4.3.3. *Algorithm 2 converges to an optimal solution in finite iterations.*

Proof. The feasibility cuts are added to the master problem (4.3.3) to remove the first-stage decisions that result in infeasible M_2. It is known from Assumption 2 that the set of feasible solutions to M_1 is finite. Therefore, a finite number of inequalities of type (4.3.12) and (4.3.13) can be added to (4.3.3). Moreover, the sets Π_{rs} and $\bar{\Pi}_{rs}$ of the optimal solutions to the dual problems in Section 4.3.2 are finite since there is no constraint parallel to the objective function (4.3.14). Therefore, a finite number of optimality cuts (4.3.24) will be generated. Given that there are finite numbers of feasibility and optimality cuts, Algorithm 2 converges in a finite time following the convergence of the Benders decomposition algorithm [76]. $\square$

4.4 Numerical Experiments

Test problem instances are obtained from Leeftink and Hans [50]. The instances consist of different surgical specialties such as orthopedic, otorhinolaryngology, and oncology. The surgery durations follow a three-parameter lognormal distribution [35]. We used the Monte Carlo sampling method to generate a finite set of scenarios for random surgery durations. The overhead and variable costs for operating rooms are determined using the cost settings in [29]. The OR opening cost is calculated by multiplying the overhead cost by the OR available time. Each OR operates an 8-hour workday and has one block that is assigned to a surgical specialty. Optimization models are implemented in Python using IBM CPLEX on a workstation with 24 cores, 3 GHz processors, and 384 GB of memory. A time limit of one hour is imposed for all instances. The valid cuts are implemented using the CPLEX lazy constraint callback function.

4.4.1 Comparing CCP with Other Stochastic models

We compare the performance of the proposed chance-constrained model with the two models proposed in [67]: SDORS-EV and SDORS-CVaR. SDORS-EV is a stochastic programming model that attempts to optimize the expected value of OR overtime and patient waiting costs. SDORS-CVaR is a risk-based model that minimizes the expected tail of overtime and waiting costs by using the CVaR function [67]. For simplicity, the following terminology is used in our experiments: CCP (chance-constrained), CVaR (SDORS-CVaR) and EV (SDORS-EV). The performance of these models are evaluated using several criteria.

Table 4.3 compares the performance of the three models after solving eight OR scheduling instances within the specified time limit. A finite set of 100 scenarios is generated for each surgery, and the parameter α_r is set to 0.10. The second and third columns show the number of surgical cases and available ORs for surgery operation. The column *Time* shows the computational time in seconds. Finally, the optimality gap reported in the last column is calculated as $\left(\frac{UB-LB}{UB}\right) \times 100\%$. It can be observed that the CCP model outperforms CVaR and EV in convergence speed. The chance-constrained model can solve all instances to optimality within the specified time limit while EV and CVaR only solve instances with up to 12 and 16 patients, respectively.

Table 4.3: Computational efficiency of CCP, CVaR and EV

Instance	$\lvert I\rvert$	$\lvert R\rvert$	Model	Time (s)	Gap (%)
1	6	5	CCP	**1.6**	0.0
			CVaR	3.5	0.0
			EV	6.8	0.0
2	7	5	CCP	**2.2**	0.0
			CVaR	7.5	0.0
			EV	32.1	0.0
3	9	5	CCP	**11.3**	0.0
			CVaR	132	0.0
			EV	87.1	0.0
4	12	5	CCP	**140**	0.0
			CVaR	550	0.0
			EV	1258.8	0.0
5	16	5	CCP	**1169.5**	0.0
			CVaR	2623.8	0.0
			EV	3600	1.38
6	20	5	CCP	**1230**	0.0
			CVaR	3600	4.2
			EV	3600	5.02
7	23	10	CCP	**2204.3**	0.0
			CVaR	3600	11.7
			EV	3600	21.0
8	29	10	CCP	**2889.5**	0.0
			CVaR	3600	39.5
			EV	3600	41.0

In Figure 4.1, the trade-off between minimizing costs and controlling the variability of costs is compared for each model. Several values are used for the confidence level parameter α in order to mimic the behavior of these models under different risk attitudes. A high α resembles an aggressive approach to minimize expected costs while accepting a substantial risk of OR overtime. On the contrary, a low α depicts conservative decision-making (i.e., accepting higher costs given that the chance of OR overtime is low). As observed in Figure 4.1, CVaR places emphasis on minimizing variability while EV focuses on providing the minimum average costs. However, CCP provides a more moderate trade-off between minimizing average costs and reducing variability. Assuming a given tolerance for the OR schedule variability, CCP outperforms CVaR by providing more cost-effective solutions. Similarly, CCP outperforms EV by providing OR schedules with lower variability, assuming a fixed budget.

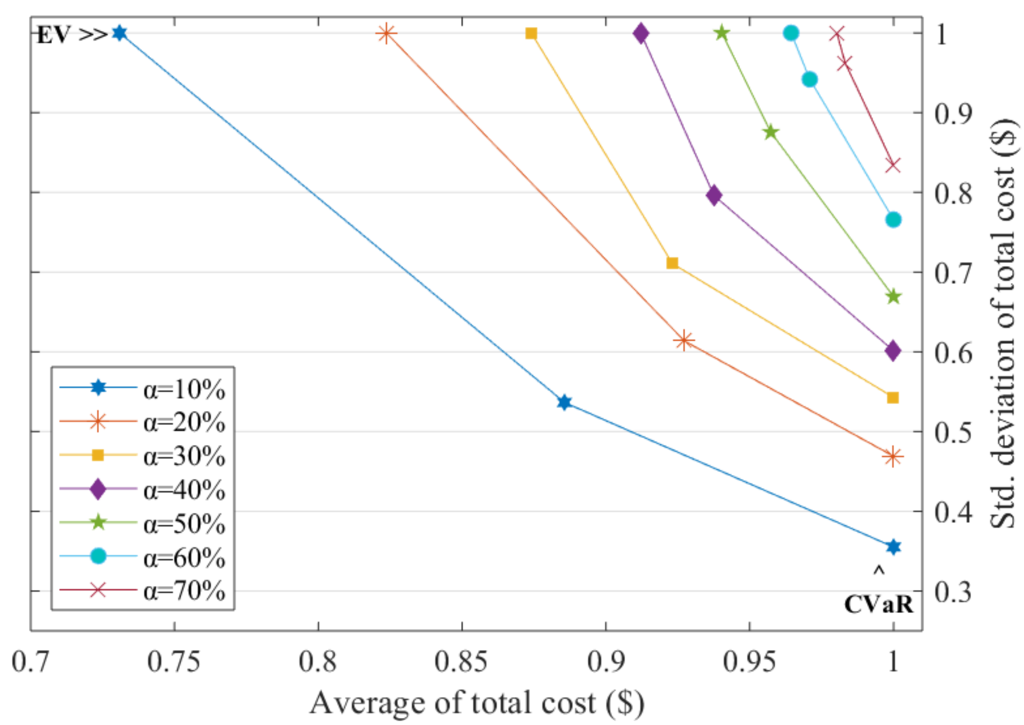
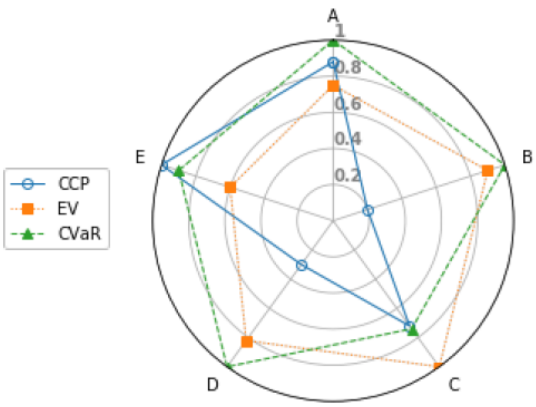


Figure 4.1: Trade-off between average total cost and variability of total cost

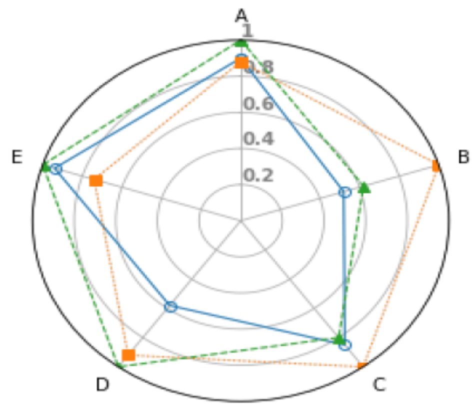
In Figure 4.2, we use several metrics to compare the performance of CCP, EV and CVaR under different values for α. The metrics are $[A, B, C, D, E]$ = [total cost, total waiting time

& overtime, utilization, overtime scenarios, open ORs].

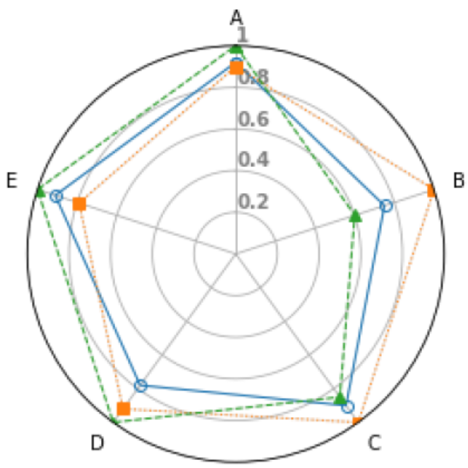
For small α values, CCP and CVaR suggest opening more ORs to reduce the risk of overtime and reduce the expected tail costs, respectively. Therefore, they incur higher average total costs and lower OR utilization than EV. EV displays better OR utilization at the risk of experiencing increased overtime. CPP is the superior method in terms of reducing overtime and patient waiting times. Moreover, CCP performs best in reducing the number of scenarios where overtime occurs. Overall, using CCP results in fewer occurrences of overtime and better OR utilization than CVaR when α is not very restrictive (i.e., $\alpha > 0.1$ in our numerical experiments). It is also observed that the three models converge in all metrics as α increases.



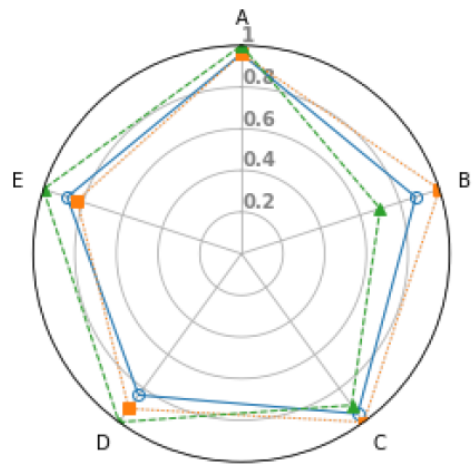
(a) $\alpha = 0.1$



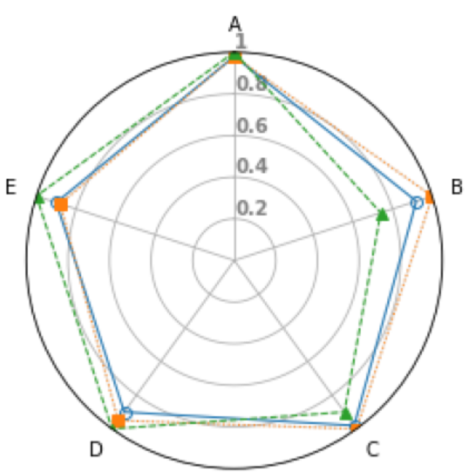
(b) $\alpha = 0.3$



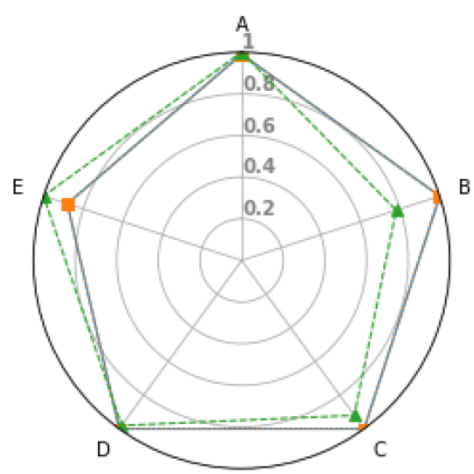
(c) $\alpha = 0.5$



(d) $\alpha = 0.6$



(e) $\alpha = 0.7$



(f) $\alpha = 0.8$

Figure 4.2: Impact of using different risk thresholds on the performance of CCP, EV and CVaR

Proposition 6. *CCP, CVaR and EV provide the same optimal solution when* $\alpha = 1$*.*

Proof. When $\alpha = 1$, the chance constraints (4.2.12) can be written as

$$Pr\{tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r : \forall k, s\} \geq 0 \quad \forall r \in R \tag{4.4.1}$$

which holds for all feasible solutions to the first-stage problem. Therefore, the chance constraints are redundant and we have $CCP \equiv EV$ when $\alpha = 1$. Now, it suffices to show that $CVaR \equiv EV$. CVaR is defined as the expectation of those outcomes where total costs exceed a threshold value, called Value-at-Risk (VaR) [78]. For $\alpha = 1$, the VaR of second-stage costs is defined as

$$VaR_1 = Min\{g(x,s) : CDF\,(g(x,s)) \geq 0\} \tag{4.4.2}$$

where CDF represents the cumulative density function. Given that $CDF \geq 0$ for every random variable, we conclude

$$g(x,s) \geq VaR_1 \quad \forall x \in X, s \in S. \tag{4.4.3}$$

Therefore, from Assumption 2 and inequality (4.4.3), the CVaR model minimizes the total costs over all scenarios, thus indicating equivalence to using the EV model. □

4.4.2 Solving Large-scale Test Instances

It is observed in Table 4.3 that the solution times increase exponentially as the problem size grows. Therefore, we apply the valid inequalities and the decomposition algorithm presented in Section 4.3 to solve larger test instances in shorter time periods. We used the Monte Carlo sampling method to generate a set of 100 scenarios, and the parameter α_r was set to 0.10. Table 4.4 compares the performance of feasibility cuts (4.3.12) and (4.3.13) when used separately and combined to solve the test problem instances shown in Table 4.3. It can be observed that both valid inequalities are effective in reducing the solution time when compared to the cuts generated by the CPLEX solver. It is also observed that using both types of cuts leads to longer solution times for small test instances due to the time spent for generating inequalities. However, as problem size grows, adding both types of feasibility cuts to $\tilde{\mathbb{F}}$ leads to significantly faster convergence than applying them separately. Therefore, we use valid inequalities (4.3.12) and (4.3.13) to generate feasibility cuts in the following numerical experiments.

Larger test problem instances are solved and reported in Table 4.5 to evaluate the performance of the proposed decomposition algorithm. We compare the performance of our algorithm with that of the IBM CPLEX MIP Solver 12.9 and a decomposition algorithm that uses the big-M optimality cuts introduced in [54]. The solution time and the optimality gap are reported for each algorithm. The column *Basic Decomposition* illustrates the results from the decomposition algorithm using feasibility cuts (4.3.12) and big-M optimality cuts. The last column highlights the results of the proposed decomposition algorithm in this chapter. As shown in Table 4.5, we observe that the CPLEX solver is the least desirable option for solving M_{DEF}, as expected. For the largest problem instance, the CPLEX solver does not find any feasible solutions within the time limit. User-defined feasibility and optimality cuts can improve solution speed. It can also be observed that using stronger optimality cuts rather than the big-M cuts can reduce solution time significantly. Neither the CPLEX solver nor the basic decomposition algorithm can solve any of the instances to global optimality within the time limit. Nevertheless, the proposed decomposition algorithm outperforms other methods by solving all test instances to optimality within 48 minutes.

Table 4.4: Performance of different feasibility cuts

Instance	Surgeries	ORs	**Cuts** (4.3.12)		**Cuts** (4.3.13)		**Both Cuts**	
			Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
1	6	5	2.5	0.0	1.9	0.0	3	0.0
2	7	5	2.6	0.0	2	0.0	5.1	0.0
3	9	5	8.2	0.0	7.5	0.0	9.8	0.0
4	12	5	69.1	0.0	75	0.0	49.6	0.0
5	16	5	205.6	0.0	243.7	0.0	131.3	0.0
6	20	5	216.2	0.0	285.2	0.0	150.5	0.0
7	23	10	357.9	0.0	405.1	0.0	231.7	0.0
8	29	10	653.4	0.0	732.9	0.0	405.4	0.0

Table 4.5: Performance of different solvers/algorithms on large-scale problems

Instance	Surgeries	ORs	Scenarios	CPLEX		Basic Decomposition		This Chapter	
				Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
9	40	10	100	3600	2.3	3600	1.5	713.8	0.0
	40	10	500	3600	5.6	3600	4.0	1400.3	0.0
10	63	20	100	3600	11.2	3600	9.8	1333.2	0.0
	63	20	500	3600	16.8	3600	13.4	2246.6	0.0
11	74	20	100	3600	19.4	3600	15.9	2270.8	0.0
	74	20	500	3600	28.7	3600	21.7	2631.6	0.0
12	89	20	100	3600	45	3600	32.3	2819.5	0.0
	89	20	500	-	-	3600	51.5	2851.2	0.0

4.4.3 Importance of Minimizing Expected Costs

Our numerical experiments show that a significant portion of total costs comes from the expected overtime and waiting time costs. Neglecting these measures in OR scheduling models can result in surpassing the predicted overtime budget by 200%, disheartening staff from longer-than-expected shifts, and causing dissatisfaction to patients [31]. We solved 10 replications of all test instances in Table 4.3 using two different objective functions: objective (4.2.22); and objective (4.2.22) minus the expected second-stage costs. Then, we calculated the sum of costs pertaining to OR opening, OR overtime, and patient waiting time for each optimal solution of each case. The percentage of savings obtained from optimizing both deterministic and stochastic costs is calculated as $\left(\frac{obj_2^*-obj_1^*}{obj_2^*} \times 100\%\right)$. It is observed from Figure 4.3 that minimizing the expected costs leads to greater savings when larger α values are used. This highlights the importance of minimizing the expected costs in addition to satisfying the chance constraints when solving stochastic OR scheduling problems.

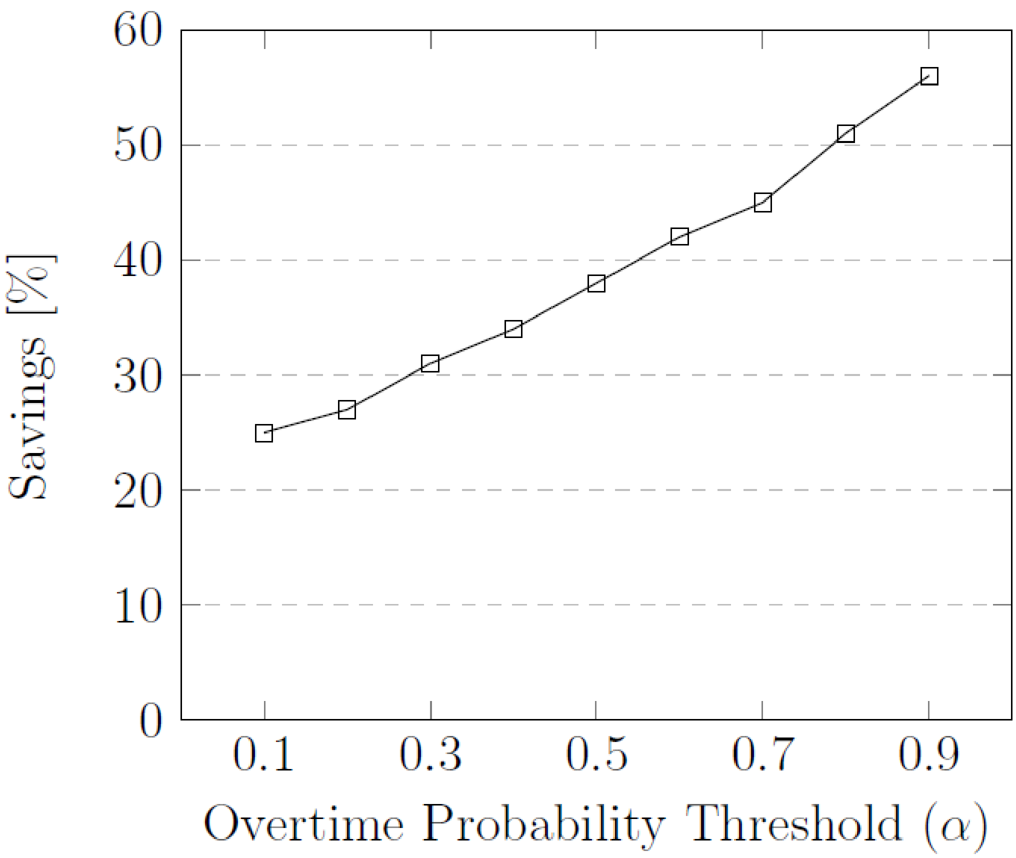


Figure 4.3: Advantage of minimizing expected costs when solving the chance-constrained OR scheduling model

4.4.4 Individual vs. Joint Chance Constraints

The chance constraints in Section 4.2.2 are enforced on each OR independently. However, a decision-maker might be interested in controlling the chance of OR overtime on an aggregate level. In such cases, the chance constraints (4.2.12) are replaced by

$$Pr\{tr_{krs} + \sum_{i \in I} \delta_{is} y_{ikr} \leq cap_r, \forall k, r, s\} \geq 1 - \alpha. \tag{4.4.4}$$

This section compares the joint chance-constrained OR scheduling model (M_{Joint}) with the proposed model M_{DEF} presented in Section 4.2. In the following numerical experiments, the Monte Carlo sampling method is used to generate a set of 100 scenarios for each test instance. We applied the decomposition algorithm proposed in Section 4.3 and the same classes of valid inequalities to both models. Figures 4.4 and 4.5 compare M_{DEF} and M_{Joint} after solving the larger test instances shown in Table 4.5. Figure 4.4 illustrates that M_{Joint} tends to open more ORs to satisfy the tighter limit on OR overtime. The joint chance constraints restrict the occurrence of overtime to $\alpha|S|$ scenarios while the individual chance constraints allow up to $\min\{\alpha_r|S||R|, |S|\}$ scenarios with OR overtime. As the probability threshold α loosens, the gap between the optimal number of open ORs obtained by the two models shrinks due to the converging feasible regions. Similar to Proposition 6, it can be shown that the two models achieve equivalent optimal solutions when $\alpha = 1$. Figure 4.5 compares the performance of M_{Joint} and M_{DEF} in reducing the OR overtime and patient waiting times. The vertical axes depict the average overtime and waiting time per OR per scenario, respectively. It can be observed that M_{Joint} achieves greater success at controlling OR overtime by opening more ORs and setting the projected start times earlier to satisfy the stricter chance constraints. However, these measures lead to lower OR utilization and

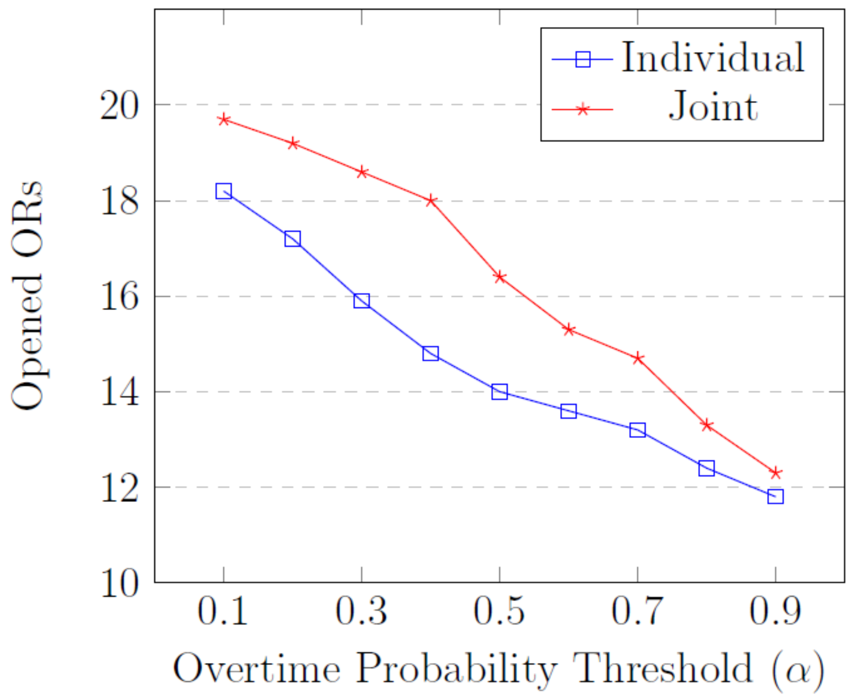


Figure 4.4: Optimal number of open ORs

higher waiting times when compared to that of M_{DEF}.

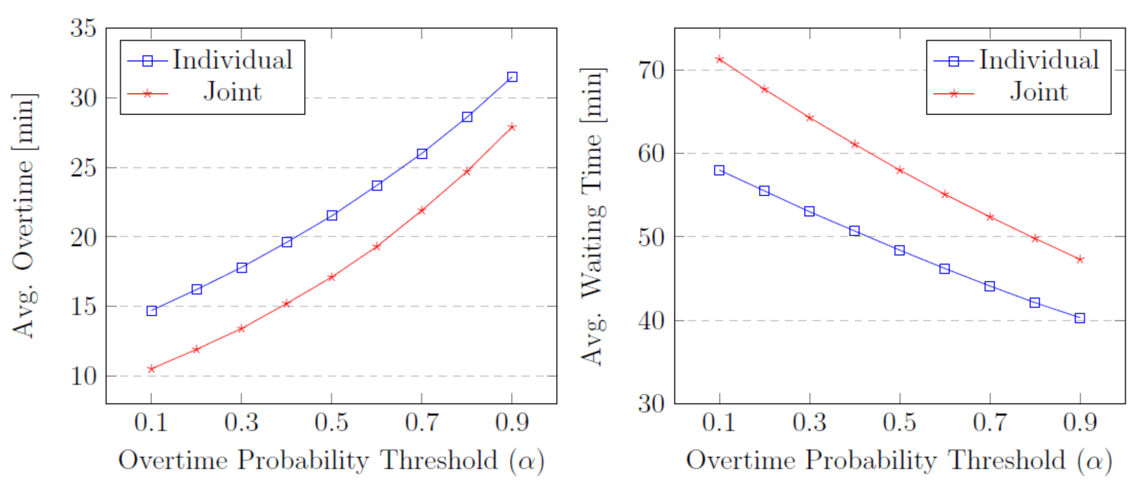


Figure 4.5: Optimal overtime and waiting times

Table 4.6 compares the run times of the two models after solving the test instances presented in Table 4.3 and Table 4.5. Column *First Feasible* shows the run time to find the first feasible solution. Column *1% Gap* details the run time to reach 1% optimality gap. It is observed that M_{DEF} can find feasible solutions and attain near-optimal solutions in shorter time lengths compared to M_{Joint}. For the largest test instance, M_{DEF} finds a feasible solution within half the time required by M_{Joint} and solves the problem to optimality within 48 minutes. On the other hand, M_{Joint} fails to reach 1% optimality gap within the one-hour time limit.

Table 4.6: Computational performance of M_{Joint} and M_{DEF}

Instance	M_{Joint}		M_{DEF}	
	First Feasible (s)	1% Gap (s)	First Feasible (s)	1% Gap (s)
1	1.2	4.7	0.8	3.1
2	1.7	7.8	1.3	5.5
3	2.6	13.8	1.8	9.5
4	7	50.1	5	38.8
5	26.6	151	19.7	111.1
6	23.5	163.7	16.6	121.9
7	50.7	244.4	34.8	179.3
8	73.8	448.3	52.7	317.4
9	137.1	988.9	80.2	649.3
10	296.8	2066	172.5	1312.2
11	519.1	3056.2	339.3	2090.7
12	616.4	3609.9	316.8	2594.6

4.5 Conclusions

In this chapter, a chance-constrained mixed-integer programming model was proposed for the OR scheduling problem with stochastic surgery durations. The individual chance constraints controlled the risk of OR overtime. The goal was to minimize the sum of OR opening, OR overtime and patient waiting costs. Our model was compared with two other stochastic models in the literature: an expected value model and a CVaR-based model. We demonstrated that minimizing the expected costs when solving the chance-constrained OR scheduling model results in significant savings compared to the case where only the deterministic costs are minimized. Moreover, we compared the individual and joint chance constraints in terms of allocated ORs, second-stage stochastic costs and solution times. A decomposition algorithm with strong feasibility and optimality cuts was applied to effectively solve large-scale test instances. We proposed an algorithm that generated feasibility cuts using the first stage solutions, and as a result, significantly reduced the time required to find feasible solutions. Numerical experiments demonstrated that the decomposition algorithm outperformed both the IBM CPLEX solver and a basic decomposition algorithm by solving the largest test instances to optimality within the one-hour time limit. Moreover, it is shown that the individual chance constraints lead to higher OR utilization, reduced patient waiting times and shorter solution times. This work has the following limitations. First, it is assumed that the surgery duration follows a known probability distribution. Second, we considered the upstream and downstream resources (e.g., nurses, beds) to be sufficiently available. However, such assumptions do not hold true for all real-life cases. It is also demonstrated that finding strong cuts can increase the convergence speed significantly. Therefore, relaxing the above assumptions and discovering stronger feasibility and optimality cuts in order to solve more complex problems in shorter time frames can be a promising topic for future research.

Chapter 5

Using Lagrangian Relaxation and Conditional Value-at-Risk Approximations to Develop an Augmented Decomposition Algorithm for the Stochastic Surgery Scheduling Problem

5.1 Introduction

Operating rooms contribute to more than 30% of total expenses [56] and 40% of total revenues [33] in hospitals. The rapid growth of healthcare expenditures adds to the significance of optimal operating room management. This increase in expenditures is fueled by new technologies, new medications, aging population and the shortage of skilled staff in hospitals [51]. Surgical expenses contribute to 30% of healthcare expenditures. Hospitalizations involving surgical procedures constituted 29% of total hospital stays while contributing to 48% of the total hospital costs in the US [94]. In light of these reports, surgeries are recognized amongst the most crucial activities in hospitals from social, medical and economic points of view. The OR scheduling problem can be categorized into four planning levels: case mix planning, master surgery scheduling, daily surgery scheduling and online monitoring/re-scheduling [12]. The daily surgery scheduling problem itself can be divided into two stages. The first stage addresses the assignment of patients to specific days over a given planning horizon. In the second stage, the scheduled patients are sequenced and surgery start times are generated. In this chapter, we focus on the second stage of the daily surgery scheduling problem, that is developing a single-day schedule for a set of elective operations in an operating theater. The schedule provides the assignment of patients to operating rooms along with a projected start time for each operation.

Several survey articles have studied the OR planning and scheduling problems and identified main challenges over the years [13]. Samudra, Van Riet, Demeulemeester, Cardoen, Vansteenkiste and Rademakers [79] classified the literature using seven different perspectives on OR scheduling problems, including *uncertainty*. They found uncertain surgery durations and random emergency patient arrivals as the most commonly studied sources of uncertainty by the Operations Research community. The required time for surgical interventions may vary significantly based on the type of operations being performed, the surgical team and the patient. In another survey paper, May, Spangler, Strum and Vargas [61] recognized high variability of surgery durations as the biggest challenge towards developing practical and cost-effective OR schedules amongst other sources of uncertainty. Due to such uncertainty in scheduling, reducing variability of the performance measures in the provided schedules can improve capacity utilization, cost-efficiency and patient satisfaction [62]. Hence, it is crucial to ensure that the provided schedule works reliably in the presence of large variability in surgery durations. Stochastic programming models have been widely used in the literature to address the OR scheduling problems with uncertain surgery durations [5, 29, 59]. One of the biggest disadvantages of the stochastic programming models is that they are not scalable. In many cases, commercial solvers fail to find even feasible solutions within a reasonable solution time. A large body of heuristic and metaheuristic methods in the literature aim to solve the optimization model fast, but they do not guarantee optimality [65, 91, 95]. This chapter proposes a new bounding approach to improve the quality of initial feasible solutions found for the two-stage stochastic OR scheduling model. The strong bounds are utilized to develop an augmented decomposition algorithm that can optimally solve the stochastic model in finite iterations.

Stochastic programming models have been widely used in the literature to address the

OR scheduling problems with uncertain surgery duration [29, 59]. The majority of the existing stochastic programming models attempted to optimize the expected value of a performance measure. The concept of using the expected value (EV) in scheduling optimization can be helpful for problems with predictable variability (i.e., low risk) on the parameter. However, for problems with frequent changes in a less predictable manner and in short-term (i.e., high-risk), the optimal solution may show poor performance for specific realizations of the random parameter [67].

The chance-constrained programming (CCP) models [15] have been used by researchers to address the optimization problems under uncertainty in healthcare [42, 97]. This approach controls the risk of undesirable events (e.g., OR overtime, patient waiting time) exceeding a specified tolerance, rather than merely minimizing their expected value. Shylo, Prokopyev and Schaefer [84] applied chance-constraints to control the OR block overtime in the OR surgery planning problem. Zhang, Denton and Xie [98] studied a chance-constrained OR surgery allocation problem.

Jebali and Diabat [40] studied the surgery planning problem with uncertain surgery duration, the length of stay in Intensive Care Unit (ICU), and emergency patient arrival. Wang, Li and Peng [92] proposed a distributionally robust chance-constrained model for the surgery planning problem with stochastic surgery durations. Noorizadegan and Seifi [71] proposed a chance-constrained programming model for the surgery planning problem with uncertain surgery durations. Deng and Shen [19] developed a two-stage stochastic model for the OR scheduling problem with a joint chance constraint on OR overtime.

It is shown that the surgery scheduling problem is NP-hard [11]. A major drawback of using the stochastic programming models is that they are not scalable. In many cases, commercial solvers fail to find even feasible solutions within a reasonable solution time. A large body of heuristic and metaheuristic methods in the literature aim to solve the optimization model fast, but they do not guarantee optimality [65, 91, 95]. Moreover, generic decomposition methods such as the Benders decomposition may show poor performance in solving large-scale test instances [54, 68].

In this chapter, we develop a chance-constrained programming model for the OR scheduling problem using individual chance constraints to reduce the risk of OR overtime. The main contribution of this chapter is to develop a fast decomposition algorithm that is augmented by strong lower and upper bounds using Lagrangian relaxation and CVaR approximations.

5.2 Two-Stage Stochastic Programming Models for the OR Scheduling Problem

A frequently used stochastic programming approach in the literature is the two-stage model in which the decision-maker assigns surgical cases to operating rooms in the first stage. On the day of surgery, random surgery durations affect the outcome of the first-stage decision. Hence, a recourse decision can be made in the second stage to minimize the expected costs (e.g., overtime, waiting time) as a result of the first-stage decision. The optimal policy from such a model is a single first-stage policy and a collection of recourse actions defining which second-stage decision should be taken in response to each random outcome. It is assumed that the random parameters in the second stage are either known or can be estimated by appropriate probability distributions.

The classical two-stage linear stochastic problems can be formulated as [8]

$$\begin{aligned} \underset{x\in\Re^n}{Min} \quad & g(x) = c^T x + E[Q(x,\xi)], \\ & \text{subject to} \\ & Ax \geq b, \\ \text{and} \quad & x \in X. \end{aligned} \tag{5.2.1}$$

where $Q(x,\xi)$ is the optimal value of the second-stage problem under scenario ξ given the first-stage decisions x

$$\begin{aligned} \underset{y \in \Re^m}{Min} \quad & q(\xi)^T y, \\ \text{subject to} & \\ T(\xi)x + W(\xi)y \geq h(\xi)&, \\ \text{and} \quad y \in Y&. \end{aligned} \tag{5.2.2}$$

Sets $X \subseteq \mathbb{R}^{n_1}$ and $Y \subseteq \mathbb{R}^{n_2}$ enforce the non-negativity, continuous and/or integrality constraints on the first-stage and second-stage decision variables. Vectors $b \in \mathbb{R}^{m_1}$ and $h(\xi) \in \mathbb{R}^{m_2}$ show the right-hand side parameters. Finally, $A \in \mathbb{R}^{n_1 \times m_1}$, $T(\xi) \in \mathbb{R}^{n_1 \times m_2}$ and $W(\xi) \in \mathbb{R}^{n_2 \times m_2}$ are the coefficient matrices used in the two-stage problem. It is often assumed that the random vector ξ has a finite number of possible realizations, called scenarios, say ξ_1, ξ_2, ..., $\xi_{|\mathbb{S}|}$ with respective probability masses p_1, p_2, ..., $p_{|\mathbb{S}|}$. Then the expectation in the first-stage problem's objective function can be written as $E[(Q, \xi)] = \sum_{s=1}^{|\mathbb{S}|} p_s Q(x, \xi_s)$.

The two-stage problem can be formulated as one large mixed-integer programming model, called the deterministic equivalent formulation (DEF) of the original problem

$$\begin{aligned} Min \qquad & f^T x + p_1 q(1)^T y(1) + p_2 q(2)^T y(2) + ... + p_{|\mathbb{S}|} q(|\mathbb{S}|)^T y(|\mathbb{S}|), \\ \text{subject to} & \\ & T(1)x + W(1)y(1) \geq h(1), \\ & T(2)x + W(2)y(2) \geq h(2), \\ & \quad . \quad . \quad . \\ & \quad . \quad . \quad . \\ & \quad . \quad . \quad . \\ & T(|\mathbb{S}|)x + W(|\mathbb{S}|)y(|\mathbb{S}|) \geq h(|\mathbb{S}|), \\ & \text{and} \quad x \in X, y(s) \in Y, s \in \mathbb{S}. \end{aligned} \tag{5.2.3}$$

Introducing a large number of scenario-specific variables, i.e., y_s, combined with the existing integrality constraints adds to the computational complexity of the resulting model. Solving DEF in its current form is a very challenging task for commercial solvers.

Decomposition algorithms have been widely used in the literature to solve large-scale mixed-integer programming models [19, 68, 76]. Lagrangian methods were applied in the early 1970s to general integer programming problem [28, 82] and scheduling problems [26]. Several papers in the literature have used Lagrangian relaxation to solve the OR scheduling problem [3, 32, 99]. This chapter proposes a scalable decomposition algorithm to solve large-scale instances of the two-stage chance-constrained OR scheduling problems efficiently. The Lagrangian Relaxation and CVaR methods are utilized to derive lower and upper bounds for the proposed model, respectively.

5.2.1 Mathematical Formulation

Let I be the set of elective surgeries and R be the set of operating rooms. The problem is to schedule surgeries over a daily planning horizon. We assume that operating rooms use the block booking policy [25]. Each OR is to be assigned to a surgical team or a surgeon via the master surgery schedule. An incidence matrix $E = \{e_{ir}\}$ for all $i \in I, r \in R$ is constructed to show the eligible surgery-to-OR assignments. The surgery duration is random, denoted by a random vector $\delta = (\delta_1, ..., \delta_{|I|})^T \in \mathbb{R}_+^{|I|}$, where δ_i shows the random duration for surgery $i \in I$. We assume that the random surgery duration has finite and discrete support $\mathbb{S}$ for δ. The probability density of each scenario s is denoted by p_s, where $\sum_{s\in\mathbb{S}} p_s = 1$. Each realization of δ in scenario s is shown by $\delta^s = (\delta_1^s, ..., \delta_{|I|}^s)^T$. Every surgery on the daily booking list must be operated. No interruption is allowed once an operation has started. It is desired to avoid OR overtime with high probability. This restriction is enforced using chance constraints. Each surgery must be assigned an OR and a projected start time. Our goal is to minimize the total costs corresponding to OR opening and overtimes, and patient waiting times.

5.2.2 Chance-Constrained OR Scheduling Problem

We propose a two-stage stochastic model for the daily OR scheduling problem with individual chance constraints enforcing overtime restrictions on every OR. Table 5.1 introduces the notation used in our model:

Table 5.1: Sets, parameters and variables used in the model

	Symbol	Definition
Sets	I	elective surgery, $i \in \{1, ..., \|I\|\}$
	R	operating room, $r =\in \{1, ..., \|R\|\}$
	K	order of surgery appointment in OR, $k \in \{1, ..., \|K\|\}$
	$\mathbb{S}$	scenario, $s \in \{1, ..., \|\mathbb{S}\|\}$
Parameters	e_{ir}	=1 if surgery i can be assigned to OR r; 0 otherwise
	δ_{is}	duration of surgery i in scenario s
	f_r	fixed cost of opening OR r
	cap_r	operating time limit for OR r
	co_r	unit overtime cost of OR r
	cw_r	unit waiting cost for surgery i
	p_s	probability density of scenario s
	α_r	overtime probability threshold for OR r (confidence level), $\alpha_r \in (0, 1)$
	M	a sufficiently large number
Variables	u_r	=1 if OR r is open; 0 otherwise
	y_{ikr}	=1 if surgery i scheduled as k^{th} surgery in OR r; 0 otherwise
	tp_{kr}	projected start time for surgery i
	tr_{krs}	actual start time for surgery i in scenario s
	o_{rs}	OR r overtime in scenario s
	w_{rs}	total patient waiting times in OR r and scenario s
	z_{rs}	=1 if chance constraint on OR r is violated in scenario s, 0 otherwise

The first-stage problem involves deterministic decision-making (i.e., OR opening and surgery case assignment) prior to the realization of uncertain surgery durations. The first-stage problem (M_1) can be formulated as

$$\begin{aligned} & Min \sum_{r \in R} f_r u_r \\ & \text{and} \quad x \in X \end{aligned} \tag{5.2.4}$$

where $x = (u, y)$ is the vector of first-stage variables. Set X is the resulting set from the deterministic constraints (5.2.5)-(5.2.9) formulated as

$$y_{ikr} \leq u_r \qquad \forall i \in I, k \in K, r \in R, \tag{5.2.5}$$

$$\sum_{k,r} y_{ikr} = 1 \qquad \forall i \in I, \tag{5.2.6}$$

$$y_{ikr} \leq e_{ir} \qquad \forall i \in I, k \in K, r \in R, \tag{5.2.7}$$

$$\sum_{i} y_{ikr} \geq \sum_{i} y_{i(k+1)r} \qquad \forall k \in K \setminus \{|K|\}, r \in R, \tag{5.2.8}$$

$$\text{and} \quad u_r \in \mathbb{B}^{|R|}, y_{ikr} \in \mathbb{B}^{|I| \times |K| \times |R|} \qquad \forall i \in I, k \in K, r \in R. \tag{5.2.9}$$

Objective function (5.2.4) minimizes the total cost of opening operating rooms. Constraints (5.2.5) and (5.2.6) ensure that every surgery will be assigned to one and only one spot in an open OR during the day. Constraint (5.2.7) enforces eligible surgery-to-OR assignments. Constraint (5.2.8) determines the order of operating surgical cases in each OR. Constraint (5.2.9) enforces binary values for the first-stage decision variables.

After the uncertain parameters are revealed, the second-stage problem determines recourse actions (e.g., adjusting start times and adding OR overtime) that incur additional costs to provide meaningful schedules based on the first-stage decision. The goal is to minimize the total cost as well as satisfying the chance constraints on OR overtime. The chance-constrained second-stage problem (M_2) is formulated as

$$Min \quad \mathbb{E}_s \left[g(x, s) \right] =$$

$$\mathbb{E}_s\left(\sum_r (1-z_{rs})g_r^1(x,s) + \sum_r z_{rs}g_r^2(x,s)\right), \tag{5.2.10}$$

$$tp_{kr} \leq tp_{(k+1)r} \qquad \forall k \in K \setminus \{|K|\}, r \in R, \tag{5.2.11}$$

$$tr_{krs} \geq tp_{kr} \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.12}$$

$$tr_{krs} \leq tr_{(k+1)rs} \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.13}$$

$$\sum_k (tr_{krs} - tp_{kr}) \leq w_{rs} \qquad \forall r \in R, s \in \mathbb{S}, \tag{5.2.14}$$

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + M z_{rs} \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.15}$$

$$\sum_{s \in \mathbb{S}} z_{rs} \leq \alpha_r |\mathbb{S}| \qquad \forall r \in R, s \in \mathbb{S}, \tag{5.2.16}$$

$$tp_{kr}, tr_{krs}, o_{rs}, w_{rs} \geq 0 \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.17}$$

$$\text{and} \quad z_{rs} \in \mathbb{B}^{|R| \times |\mathbb{S}|} \qquad \forall r \in R, s \in \mathbb{S}. \tag{5.2.18}$$

Objective function (5.2.10) minimizes the expected costs corresponding to OR overtimes and patient waiting times, where

$$g_r(x,s) = \begin{cases} g_r^1(x,s) = cw_r w_{rs} & z_{rs} = 0, \quad (5.2.19) \\ g_r^2(x,s) = cw_r w_{rs} + co_r o_{rs} & z_{rs} = 1, \quad (5.2.20) \end{cases}$$

Non-negative variable o_{rs} is to capture the OR overtime. Constraint (5.2.11) is to determine the projected start time for each surgery according to the sequencing decisions. Constraint (5.2.12) ensures that each surgery starts after its projected start time. Constraint (5.2.13) is similar to (5.2.11) in that the projected start times must follow the sequencing decisions. Constraint (5.2.14) calculates the amount of waiting time in every OR per scenario. Constraints (5.2.15) and (5.2.16) enforce the chance constraints stating that the surgeries assigned to an OR must be finished during the regular hours (i.e., no overtime) with a high probability. Constraints (5.2.17) and (5.2.18) enforce non-negative and binary values for the second-stage decision variables, respectively.

Set $\mathbb{P}(s)$ of first-stage solutions satisfying the chance-constrained second-stage problem is derived as

$$\mathbb{P}(r,s) = \left\{(u,y) \in X \mid \exists tp_{kr}, tr_{krs} : tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r \right\}, \tag{5.2.21}$$

$$\mathbb{P}(s) = \bigcap_{r \in R} \mathbb{P}(r,s) = \left\{(u,y) \in X \mid \exists tr_{krs} : tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r, \forall r \in R \right\}. \tag{5.2.22}$$

The deterministic equivalent formulation for the two-stage chance-constrained OR scheduling model (M_{DEF}) can be modeled as [68]

$$M_{DEF}: \qquad Min \sum_{r \in R} f_r u_r + \frac{1}{|\mathbb{S}|} \sum_{r,s} \left(cw_r w_{rs} + co_r o_{rs} \right), \tag{5.2.23}$$

$$\textit{subject to} \quad (5.2.11) - (5.2.14), (5.2.16) - (5.2.18),$$

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + z_{rs}(M - cap_r) \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.24}$$

$$tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r + (1 - z_{rs})(M - cap_r) + o_{rs} \qquad \forall k \in K, r \in R, s \in \mathbb{S}, \tag{5.2.25}$$

$$\text{and} \quad x \in X. \tag{5.2.26}$$

Suppose δ contains d independent random components, each of which has three possible realizations (e.g., low, medium and high), then the total number of scenarios is $K = 3^d$. This is a very large scale mixed-integer program that challenges even the state-of-the-art optimization solvers [46]. Moreover, using a big M value can lead to weak LP relaxations. Assigning a smaller value for M can help tighten the feasible region for the LP relaxation of M_{DEF}. Therefore, instead of setting a single large value for M, constraint-specific formulae used to calculate the big M values are developed for constraints (5.2.24) and (5.2.25) as

$$M_{rs} = \sum_i e_{ir} \delta_{is}, \forall r \in R, s \in \mathbb{S}. \tag{5.2.27}$$

The values in (5.2.27) are valid because the total operation time in each OR does not exceed the duration of all surgical cases that can be allocated to the specific operating room. The next section introduces a decomposition algorithm to address this issue and solve M_{DEF}

efficiently.

5.3 Solution Approach

This section describes a decomposition algorithm to solve the proposed model in Section 5.2.1. The proposed algorithm can solve model M_{DEF} to optimality if the following assumptions are satisfied [68]:

- The random vector $\mathbb{S}$ has discrete and finite support. Specifically, $p_s = \frac{1}{|\mathbb{S}|}$ for $s \in \mathbb{S}$.

  We have stated this assumption in the problem description in Section 2.1.

- Sets X and $\mathbb{P}(r, s)$, $s \in \mathbb{S}$ are non-empty compact sets.

  Without loss of generality, we can assume that for every $s \in \mathbb{S}$, there exists a feasible first-stage solution that satisfies the chance constraints. Therefore, sets X and $\mathbb{P}(s)$ are finite sets of points and they are compact sets.

- Sets $conv\,(\mathbb{P}r, (s))$, $s \in \mathbb{S}$ have the same recession cone, i.e., there exists $C \subseteq \mathbb{R}^N$ such that $C = \left\{\theta \in \mathbb{R}^N | x + \lambda\theta \in \mathbb{P}(s); \forall x \in \mathbb{P}(s), \lambda \geq 0\right\}$ for all $s \in \mathbb{S}$, where $N := |R| + |I| \times |K| \times |R|$.

- There does not exist an extreme ray $\tilde{\theta}$ of conv(X) with $f^T\tilde{\theta} < 0$, i.e., the two-stage problem has a bounded optimal solution.

We begin by defining feasibility ($\mathbb{F}$) and optimality ($\mathbb{O}$) sets as

$$\mathbb{F} = \left\{x \in X, z \in \mathbb{B}^{R\times\mathbb{S}} : \sum_s z_{rs} = \alpha_r|\mathbb{S}|, r \in R, z_{rs} = 0 \Rightarrow x \in \mathbb{P}(r, s), s \in \mathbb{S}\right\}, \tag{5.3.1}$$

$$\mathbb{O} = \left\{(x, z, \rho) \in \mathbb{F} \times \mathbb{R}_+ : \rho \geq \frac{1}{|\mathbb{S}|}\sum_{r,s}(1 - z_{rs})g_r^1(x, s) + z_{rs}g_r^2(x, s)\right\}. \tag{5.3.2}$$

These sets will be approximated using feasibility and optimality cuts in the following master problem (**MP**)

$$
\begin{aligned}
&\underset{x,z,\rho}{Min} \quad f^T u + \rho, \\
&\sum_s z_{rs} = \alpha_r |\mathbb{S}| \qquad \forall r \in R, \\
&x \in X, \\
&z \in \mathbb{B}^{|R| \times |\mathbb{S}|}, \\
&\rho \geq 0, \\
&(x, z) \in \tilde{\mathbb{F}}, \\
&\text{and} \quad (x, z, \rho) \in \tilde{\mathbb{O}}.
\end{aligned} \tag{5.3.3}
$$

The sets $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{O}}$ are the outer approximations of the feasibility (F) and optimality (O) sets, respectively. Najjarbashi and Lim [68] showed that the following inequalities are valid and can be used as strong feasibility and optimality cuts to define $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{O}}$

$$\gamma x + \sum_{i=1}^{l} \left(h_{g_i}(\gamma) - h_{g_{i+1}}(\gamma) \right) z_{g_i} \geq h_{g_1}(\gamma). \tag{5.3.4}$$

$$\text{and} \quad z_{rs} - \sum_{k \in K, i \in TB} y_{ikr} \geq 1 - |TB|, r \in R, s \in \mathbb{S} \tag{5.3.5}$$

where $h_{\sigma_1} \geq h_{\sigma_2} \geq ... \geq h_{\sigma_{|\mathbb{S}|}}$, $G = \{g_1, g_2, ..., g_l\} \subseteq \{\sigma_1, \sigma_2, ..., \sigma_p\}$ and $h_{t_{l+1}} = h_{\sigma_{p+1}}$. Set TB is the subset of surgeries that lead to the violation of chance constraints when they are assigned to the same OR.

$$
\begin{aligned}
&\rho + \frac{1}{|\mathbb{S}|} \sum_{r,s \in Q} \left(-\sum_k \pi^4_{krs} cap_r - \psi^2_{rs}(\tau_{rs}) \right) z_{rs} + \frac{1}{|\mathbb{S}|} \sum_{r,s \in \mathbb{S} \setminus Q} \left(-\sum_k \bar{\pi}^4_{krs} cap_r - \psi^1_{rs}(\tau_{rs}) \right) (1 - z_{rs}) \\
&\geq \frac{1}{|\mathbb{S}|} \sum_{r,s \in Q} \left(-\sum_k \pi^4_{krs} cap_r \right) + \frac{1}{|\mathbb{S}|} \sum_{r,s \in \mathbb{S} \setminus Q} \left(-\sum_k \bar{\pi}^4_{krs} cap_r \right) + \frac{1}{|\mathbb{S}|} \sum_{i,r,s} \left(\tau_{rs} \hat{y}_{ikr} \right).
\end{aligned} \tag{5.3.6}
$$

The following decomposition algorithm is proposed to solve the two-stage chance-constrained OR scheduling problem [68]. This algorithm has a similar structure to the Benders decomposition approach [76]. Rather than traditional Benders cuts, strong feasibility cuts

(5.3.4)-(5.3.5) and optimality cuts (5.3.6) are used to enhance solution speed.

Algorithm 3 Decomposition Algorithm

Initialization: $LB := -\infty$, $UB := +\infty$
while $UB - LB > \epsilon$ **do**
    Solve master problem (5.3.3).
    if (5.3.3) infeasible **then**
        • Stop. Original problem is infeasible.
    else
        • Let $(\hat{x}, \hat{z}, \hat{\rho})$ be an optimal solution to (5.3.3).
        • $LB \leftarrow f^T \hat{u} + \hat{\rho}$.
        • Check feasibility of the second-stage problem by evaluating the inequalities (5.3.4) and (5.3.5).
        if there exists violated inequalities **then**
            • Add feasibility cuts (5.3.4) and (5.3.5) to $\tilde{\mathbb{F}}$.
        else
            • $UB \leftarrow \sum_r f_r \hat{u}_r + \frac{1}{|\mathbb{S}|} \sum_{r,s} \left[g_r^1(\hat{x}, s) + g_r^2(\hat{x}, s) \right]$
            • Add optimality cuts (5.3.6) to $\tilde{\mathbb{O}}$.
        end if
    end if
end while

Although Algorithm 3 converges to the optimal solution [68], the non-convex first-stage problem can pose a limitation to the computational efficiency of this algorithm. In the remainder of this section, we derive approximation methods to find stronger bounds for M_{DEF}.

5.3.1 Deriving Lower and Upper Bounds

Many MIP based optimization models are often difficult to solve due to a relatively small number of side constraints that significantly increase computational complexity. Dualizing the computationally difficult side constraints produces an associated Lagrangian dual formulation that is easier to solve and whose optimal value gives a dual bound on the optimal objective value of the original problem. The Lagrangian model can thus be used in place of a linear programming relaxation to provide bounds in a branch-and-bound algorithm. The

Lagrangian relaxation of the master problem (5.3.3), MP_{LR}, is formulated as

$$
\begin{aligned}
&\underset{x,z,\rho}{Min} \quad \phi = f^T u + q_r \left(\sum_s z_{rs} - \alpha_r |\mathbb{S}| \right) + \rho, \\
&x \in X, z \in \mathbb{B}^{|R| \times |\mathbb{S}|}, \rho \geq 0, \\
&(x, z) \in \tilde{\mathbb{F}}, \\
&\text{and} \quad (x, z, \rho) \in \tilde{\mathbb{O}}.
\end{aligned} \tag{5.3.7}
$$

where q_r is the Lagrangian multiplier for the chance constraints (5.2.16). There are three common approaches to solve the Lagrangian relaxation [27]: (1) the subgradient method, (2) various versions of the simplex method implemented using column generation techniques, and (3) multiplier adjustment methods. The subgradient method is widely used because it is easy to program and has worked well on many practical problems. The subgradient method for solving MP_{LR} can be described as follows:

Algorithm 4 Subgradient Method for MP_{LR}

Initialization:
• Set an upper bound ϕ^0 by finding any feasible solution to MP_{LR}.
• Set an initial vector of Lagrangian multipliers $q^0 = 0$.
• Set an initial value for the scalar θ that satisfies $0 \leq \theta \leq 2$.
Subgradient Iterations:
while $iter \leq MAX_{iter}$ **do**
  • Calculate the subgradient of ϕ^{iter} over q, $\gamma^{iter} = \left(\sum_s z_{rs}^{iter} - \alpha_r |\mathbb{S}| \right)$.
  • Calculate the stepsize parameter $t^{iter} = \theta^{iter} (\phi^0 - \phi^{iter}) / \left\| \gamma^{iter} \right\|^2$.
  • Update the Lagrangian multipliers $q^{iter+1} = max \left\{ 0, q^{iter} + t^{iter} \gamma^{iter} \right\}$.
  if $\left\| \gamma^{iter+1} - \gamma^{iter} \right\| < \varepsilon$ **then**
    • Stop. Print the optimal solution.
  end if
  if no improvement in more than C iterations **then**
    • $\theta^{iter+1} = \theta^{iter} / 2$
  end if
  $iter = iter + 1$
end while

To derive a lower bound for M_{DEF}, Algorithm 3 is called where the master problem in line 3 is replaced with MP_{LR} (5.3.7). Algorithm 4 is used to obtain the optimal solution of MP_{LR}.

It can be readily shown that the Expected Value (EV) model [67] provides a lower bound for the chance-constrained OR scheduling model M_{DEF}. The EV model is equivalent to the model M_{DEF} after relaxing the chance constraints (5.2.15) and (5.2.16), i.e., setting $\alpha_r = 0$. The next step is to derive an upper bound for the two-stage chance-constrained OR scheduling problem.

Proposition 7. *Let $F_r(tr, y, \delta) = tr_{krs} + \sum_i \delta_{is} y_{ikr}$. We can use $CVaR_{1-\alpha_r}(F_r)$ to derive an upper bound for M_{DEF} by replacing the chance constraints*

$$Prob\{F_r(tr, y, \delta) \geq cap_r\} \leq \alpha_r \quad \forall r \in R \tag{5.3.8}$$

with the following inequalities

$$CVaR_{1-\alpha_r}(F_r) \leq cap_r \quad \forall r \in R. \tag{5.3.9}$$

Proof. For the probability threshold parameter α, VaR is defined as [78]

$$VaR_{1-\alpha} = \min\{\beta : Prob\{F_r(tr, y, \delta) \leq \beta\} \geq 1 - \alpha_r\}. \tag{5.3.10}$$

We show that $CVaR_{1-\alpha_r}(F_r)$ provides a convex conservative approximation for the chance constraint $Prob\{F(tr, y, \delta) \geq cap_r\} \leq \alpha_r$. From Nemirovski and Shapiro [69], we know that

$$VaR_{1-\alpha_r}(F_r(tr, y, \delta)) \leq cap_r \equiv Prob\{F(tr, y, \delta) \geq cap_r\} \leq \alpha_r \quad \forall r \in R. \tag{5.3.11}$$

The conditional value-at-risk (CVaR) of $F_r(tr, y, \delta)$ is defined as the conditional expectation of F_r when F_r is equal or greater than VaR

$$CVaR_{1-\alpha_r}(F_r) = VaR_{1-\alpha_r}(F_r) + \mathbb{E}\left([F_r - VaR_{1-\alpha_r}(F_r)]_+\right) \geq VaR_{1-\alpha_r}(F_r) \quad \forall r \in R. \tag{5.3.12}$$

From (5.3.11) and (5.3.12) and the convexity of (5.3.11), it is concluded that the inequalities (5.3.9) provide conservative convex approximations for the chance constraints in the two-stage OR scheduling problem. □

Therefore, the chance constraints (5.2.15) and (5.2.16) can be replaced by [67]

$$\epsilon_r + \gamma_{rs} \geq tr_{krs} + \sum_i \delta_{is} y_{ikr} \qquad \forall k \in \mathbb{K}, r \in R, s \in \mathbb{S}, \tag{5.3.13}$$

$$\epsilon_r + \frac{1}{\alpha_r} \sum_s \gamma_{rs} \leq cap_r \qquad \forall r \in R, \tag{5.3.14}$$

$$\text{and} \quad \gamma_{rs} \geq 0 \qquad \forall r \in R, s \in \mathbb{S}. \tag{5.3.15}$$

where ϵ_r represents $VaR_{1-\alpha_r}$ and $\gamma_{rs} = [F_r(tr, y, \delta_s) - \epsilon_r]_+$.

Replacing the chance constraints with the conservative CVaR approximation converts the original problem to a two-stage stochastic model with relatively complete recourse, i.e., the second-stage problem is feasible for every $x \in X$. The resulting model can be solved using the Benders decomposition algorithm effectively [46]. The dual problem of the linear second-stage problem for each OR and random scenario is formulated to derive the Benders optimality cuts as follows:

$$\varrho_{rs}(\hat{x}) =$$

$$\underset{\pi}{Max} \sum_k \pi^4_{krs} \left(\sum_i \tilde{d_{is}} \hat{y}_{ikr} - cap_r \right) + \sum_k \pi^6_{krs} \left(\sum_i \tilde{d_{is}} \hat{y}_{ikr} \right) - \pi^7_r cap_r, \tag{5.3.16}$$

$$\left(\pi^1_{(k-1)rs} - \pi^1_{krs}\right) + \pi^5_{rs} \leq 0 \qquad \forall k \in \mathbb{K}, \tag{5.3.17}$$

$$\left(\pi^2_{(k-1)rs} - \pi^2_{krs}\right) + \pi^3_{krs} - \pi^4_{krs} - \pi^5_{rs} - \pi^6_{krs} \leq 0 \qquad \forall k \in \mathbb{K}, \tag{5.3.18}$$

$$\pi^4_{krs} \leq co_r \qquad \forall k \in \mathbb{K}, \tag{5.3.19}$$

$$\pi^5_{rs} \leq cw_r, \tag{5.3.20}$$

$$\pi^6_{krs} - \pi^7_r = 0 \qquad \forall k \in \mathbb{K}, \tag{5.3.21}$$

$$\pi^6_{krs} - \frac{1}{\alpha_r} \pi^7_r \leq 0 \qquad \forall k \in \mathbb{K}, \tag{5.3.22}$$

$$\sum_k \left(\pi^1_{krs} + \pi^2_{krs} + \pi^3_{krs} + \pi^4_{krs} + \pi^6_{krs}\right) + \pi^5_{rs} + \pi^7_r = 1, \tag{5.3.23}$$

$$\text{and} \quad \pi^1_{krs}, \pi^2_{krs}, \pi^3_{krs}, \pi^4_{krs}, \pi^5_{rs} \geq 0 \qquad \forall k \in \mathbb{K}. \tag{5.3.24}$$

After solving the above problem to optimality, the Benders optimality cuts can be formulated as

$$\rho \geq \frac{1}{|\mathbb{S}|} \sum_{r,s} \left[\sum_{k} \pi^{4}_{krs} \left(\sum_{i} \tilde{d_{is}} \hat{y}_{ikr} - cap_r \right) + \sum_{k} \pi^{6}_{krs} \left(\sum_{i} \tilde{d_{is}} \hat{y}_{ikr} \right) - \pi^{7}_{r} cap_r \right]. \tag{5.3.25}$$

Algorithm 5 describes the augmented decomposition algorithm (LRCVaR) using the lower and upper bounds derived by solving the Lagrangian and CVaR approximations of M_{DEF}.

Algorithm 5 LRCVaR

Input: sets and parameters in Table 5.1, model M_{DEF}, optimality gap threshold (ϵ).
Initialization: $LB := -\infty$, $UB := +\infty$, $iter := 0$
• Solve the Lagrangian relaxation approximation of M_{DEF} and get the optimal cost LB_{LR}. Solve the EV model and get the optimal cost LB_{EV}.
• Update lower bound, $LB \leftarrow \max\{LB_{LR}, LB_{EV}\}$.
• Solve the CVaR approximation model using Benders decocmposition and get the optimal cost UB_{CVaR}.
• Update upper bound, $UB \leftarrow UB_{CVaR}$.
while $\frac{UB-LB}{UB} > \epsilon$ **do**
    if $iter = 0$ **then**
        • Set the initial feasible solution $\left(\hat{x}, \hat{z}, \hat{tp}, \hat{tr}\right)$ to the optimal solution of the CVaR approximation model.
    else
        • Solve master problem (5.3.3). Let $(\hat{x}, \hat{z}, \hat{\rho})$ be an optimal solution to (5.3.3).
    end if
    • Check feasibility of the second-stage problem by evaluating the inequalities (5.3.4) and (5.3.5).
    if there exists violated inequalities **then**
        • Add feasibility cuts (5.3.4) and (5.3.5) to $\tilde{\mathbb{F}}$.
    else
        $LB \leftarrow \max\left\{LB, f^T\hat{u} + \hat{\rho}\right\}$.
        $UB \leftarrow \min\left\{UB, \sum_r f_r\hat{u}_r + \frac{1}{|\mathbb{S}|}\sum_{r,s}\left[g^1_r(\hat{x}, s) + g^2_r(\hat{x}, s)\right]\right\}$
        Add optimality cuts (5.3.6) to $\tilde{\mathbb{O}}$.
    end if
    $iter = iter + 1$.
end while
Output: optimal cost obj^* and decision variables $(x^*, z^*, tp^*, tr^*, o^*, w^*)$.

5.3.2 Selecting a Proper Size for Finite Samples

Selecting a larger sample size generally results in more accurate approximations of stochastic models [16]. However, it is shown in section 5.2.2 that the size of model M_{DEF} can grow exponentially as the number of possible values (i.e., scenarios) for the random parameters — hence the number of second-stage variables t_{krs} — increases. Therefore, it is important to take a sufficiently large sample that approximates the stochastic model accurately as well as caps the model complexity. Kleywegt et al. [44] showed that as the sample size increases, the probability that an optimal solution to the sample average approximation (SAA) model converges to the optimal solution of the original problem approaches one exponentially fast.

Algorithm 6 illustrates a SAA-based approach to find desirable sample sizes for the stochastic OR scheduling problem.

The verification sample size $N^{'}$ is set to a very large number (e.g., 10000) to approximate the original stochastic model with an accuracy close to 1. It should be noted that one can calculate $\bar{g}_m^{N^{'}}(\hat{x})$ efficiently because such task does not require any optimization.

5.4 Numerical Experiments

In this section, we evaluate the performance of the proposed decomposition algorithm in solving large-scale test instances of the stochastic OR scheduling problem. The solution speed of the algorithm is also compared with Benders decomposition algorithm [76], the strong-cut decomposition algorithm (SCDA) in [68] and the IBM CPLEX solver.

5.4.1 Test Setup

Test problem instances are obtained from Leeftink and Hans [50]. The instances are different from each other in terms of surgery types and specialties such as orthopedic, otorhinolaryngology, and oncology. All numerical experiments in this chapter are based on the following configuration. The surgery durations are assumed to follow a three-parameter lognormal distribution [35]. Scenarios are generated for random surgery duration using the

Algorithm 6 SAA-Based Method to Find Sample Size

Input: Sample size (N), Number of replications (M), Increment (c), approximation gap (μ), accuracy threshold (ϵ), verification sample size ($N^{'}$).
Initialization: $N \leftarrow 0$, $\mu \leftarrow +\infty$.
while $\mu > \epsilon$ **do**
    • $N \leftarrow N + c$.
    for replication $m = 1, ..., M$ **do**
        • Generate a sample of size N for the random surgery durations.
        • Solve the scenario-based OR scheduling model M_{DEF}. Let $\hat{obj}_m^N$ and $\hat{x}_m^N$ be the optimal objective value and optimal first-stage solutions, respectively.
        • Generate a verification sample of size $N^{'}$. Given $\hat{x}_m^N$, calculate the second-stage variables for the verification sample and estimate the optimal objective value of the original model as

$$\hat{g}_m^{N^{'}}(\hat{x}) = \frac{1}{N^{'}} \sum_{s=1}^{N^{'}} \left[\sum_r (cw_r w_{rs} + co_r o_{rs}) \right].$$

    end for
    • Calculate the average lower bound ($\bar{obj}_m^N$) and upper bound ($\bar{g}_m^{N^{'}}(\hat{x})$) for the original model [44] as

$$\bar{obj}^N = \frac{1}{M} \sum_{m=1}^{M} \hat{obj}_m^N \quad \text{and} \quad \bar{g}^{N^{'}}(\hat{x}) = \frac{1}{M} \sum_{m=1}^{M} \hat{g}_m^{N^{'}}(\hat{x}).$$

    • Calculate an upper bound of the accuracy gap μ as

$$\mu = \bar{g}^{N^{'}}(\hat{x}) - \bar{obj}^N.$$

end while
Output: Selected sample size (N^*).

Monte Carlo sampling method. The OR opening cost is calculated by multiplying the OR overhead cost into the OR available time. For simplicity, all ORs are planned to operate for 8 hours for the day. Each OR has one block that is assigned to a surgeon to perform surgical cases. Optimization models are implemented in Python using IBM CPLEX on a workstation with 24 cores, 3Ghz processors, and 384GB of memory. A time limit of 60 minutes is enforced for all instances. The valid cuts are implemented using the CPLEX lazy constraint callback function. The feasibility and optimality cuts are added to the master problem whenever CPLEX finds an integer solution to the master problems (5.3.3) or (5.3.7) that violates the inequalities (5.3.4), (5.3.5)) and (5.3.6) (or (5.3.25)), respectively.

Table 5.2 illustrates the set of 10 test problem instances solved in this chapter. The column *Surgeries* shows the number of surgical cases to schedule during the daily planning horizon. The column *OR* shows the number of available operating rooms in the operating theater. The column *CV* illustrates the average coefficient of variation for the surgical cases considered in each test instance. The size of the uncertainty set, $|\mathbb{S}|$, is shown in column *Scenarios*. The last two columns show the number of binary variables and constraints in model M_{DEF}, respectively. Each test problem is solved using two different sample sizes to examine the scalability of the solution algorithms. The overtime probability threshold, α_r, is set to 0.1 for all operating rooms. Each instance is solved for 5 replications.

Table 5.2: Test problem instances

Instance	Surgeries	OR	CV	Scenarios	Binary Variables	Constraints
1	40	10	0.617	100	17,010	211,840
2	40	10	0.617	500	21,010	863,840
3	63	20	1.055	100	81,400	750,723
4	63	20	1.055	500	89,400	2,790,723
5	74	20	0.713	100	111,540	929,594
6	74	20	0.713	500	119,540	3,321,594
7	89	20	0.605	100	160,440	1,196,909
8	89	20	0.605	500	168,440	4,068,909
9	101	25	0.585	100	257,550	1,787,726
10	101	25	0.585	500	267,550	5,857,726
11	116	25	0.957	100	338,925	2,182,616
12	116	25	0.957	500	348,925	6,852,616
13	132	30	0.575	100	525,750	3,169,212
14	132	30	0.575	500	537,750	9,541,212
15	151	40	1.205	100	916,080	5,176,351
16	151	40	1.205	500	932,080	14,888,351
17	185	40	0.771	100	1,373,040	7,093,985
18	185	40	0.771	500	1,389,040	18,981,985
19	207	40	0.621	100	1,718,000	8,482,647
20	207	40	0.621	100	1,734,000	21,778,647

5.4.2 Numerical Results

First, the SAA-based algorithm is called to determine a desirable sample size. We evaluated the accuracy of different sample sizes on the test instance 9 with 101 surgeries and 25 ORs using Algorithm 6. The accuracy threshold ϵ is set to 1% and the initial sample size $N = 10$. We increased the sample size by increments of $c = 10$ until the desired accuracy is obtained. Each iteration is solved for $M = 10$ replications and the verification sample size $N' = 10,000$. Figure 5.1 illustrates that a finite sample with 100 scenarios can achieve the desired 99% accuracy in approximating the original stochastic OR scheduling model. Increasing the sample size to 200 can increase the accuracy to 99.9%.

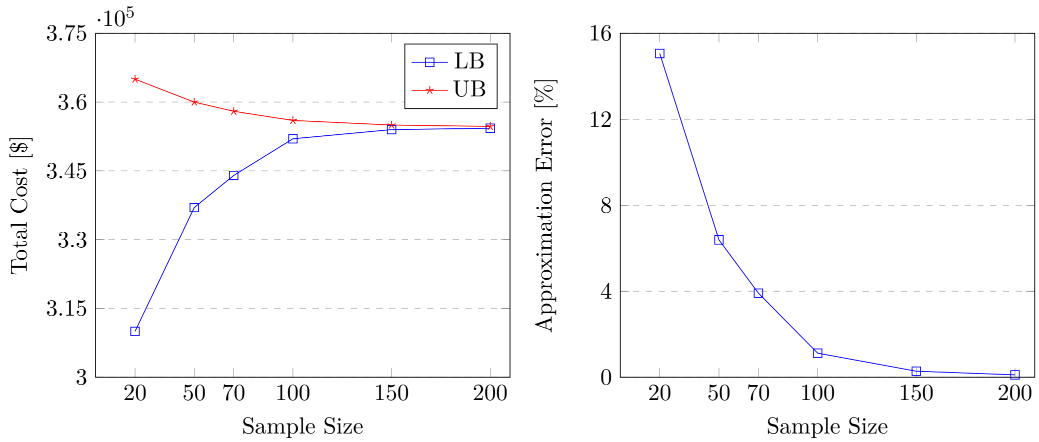


Figure 5.1: Impact of Sample Size on Approximation Accuracy

Table 5.3 and Figure 5.2 portrait the performance of each algorithm in solving the test problem instances when $\alpha_r = 0.10$. A "-" entry in Table 5.3 shows that the algorithm failed to find a feasible solution within the time limit. It can be observed that the CPLEX solver was the least desirable option for solving the test instances because it failed to find a feasible solution within the time limit when the number of surgeries and ORs exceeded 89 and 20, respectively. The Benders decomposition method outperformed the CPLEX solver, and it found a feasible solution for test instances with up to 151 surgeries and 40 ORs within one hour. However, it solved only one test problem to optimality and failed to provide near-optimal solutions when the number of surgical cases exceeded 60. The strong feasibility and optimality cuts in SCDA improved solution speed compared to the traditional Benders cuts. The SCDA solved instances with 74 surgeries to optimality and provided acceptable optimality gaps for instances of up to 132 surgeries and 30 ORs. However, as the problem size was increased and the number of surgeries to schedule went beyond 150, the optimality gap increased up to 15%. The proposed LRCVaR algorithm outperformed other methods in solution speed and quality. Despite Benders and SCDA that required more than 60 minutes for solving instances with more than 74 surgeries, LRCVaR solved instances quadrupled in size to optimality within the time limit. Moreover, it was able to solve all test instances with up to 4% optimality gap. Our decomposition algorithm achieved schedules that were at least 96.39% close to the optimal solution, for large-scale test instances with more than 200 surgeries and 40 ORs.

Table 5.3: Performance of different methods on large-scale problems

Instances	CPLEX		Benders		SCDA		LRCVaR	
	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
1	3600	6.73	2934	0.0	901.2	0.0	432	0
2	3600	9.12	3600	0.87	1893.6	0.0	562	0
3	3600	13.37	3600	1.82	1845.3	0.0	674	0.0
4	3600	16.81	3600	3.61	2934.5	0.0	741	0.0
5	3600	19.45	3600	6.92	3035.0	0.0	874	0.0
6	3600	28.77	3600	9.30	3600	0.72	1006	0.0
7	3600	45.06	3600	13.14	3600	0.89	1207	0.0
8	-	-	3600	16.89	3600	1.07	1424	0.0
9	-	-	3600	20.35	3600	1.86	1852	0.0
10	-	-	3600	22.78	3600	2.21	2166	0.0
11	-	-	3600	25.02	3600	3.01	2491	0.0
12	-	-	3600	28.64	3600	3.97	2741	0.0
13	-	-	3600	29.47	3600	4.55	3152	0.0
14	-	-	3600	33.39	3600	5.67	3472	0.0
15	-	-	3600	41.71	3600	6.12	3600	0.67
16	-	-	3600	48.20	3600	7.38	3600	1.24
17	-	-	-	-	3600	8.91	3600	1.79
18	-	-	-	-	3600	9.46	3600	2.21
19	-	-	-	-	3600	12.89	3600	2.83
20	-	-	-	-	3600	14.96	3600	3.61

5.4.3 Quality of Bounds

This section assesses the effectiveness of the proposed bounding approaches discussed in Section 5.3.1 in reducing the solution time for large-scale OR scheduling test instances. The goal is to determine if the additional time spent on solving the Lagrangian relaxation and CVaR approximations of the model M_{DEF} helps accelerate the convergence speed. A subset of the instances in Table 5.2 having 100 scenarios (i.e., odd instance numbers) is selected.

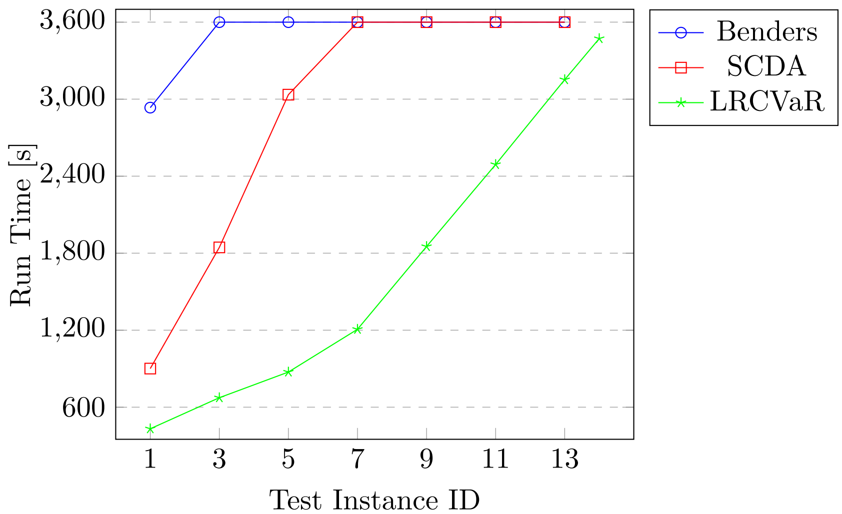


Figure 5.2: Comparing Benders, SCDA and LRCVaR in solving large-scale instances

Each instance was solved for 5 replications and α_r is set to 0.10. Table 5.4 illustrates the average run time to find the initial feasible solution and the corresponding optimality gap for Benders, SCDA and LRCVaR. The Benders algorithm failed to find a feasible solution within the time limit for instances with 150 and more surgeries. It is observed that the proposed bounds in LRCVaR enhance the performance of SCDA and achieve initial feasible solutions with 50% lower optimality gaps in shorter times. Moreover, LRCVaR achieved the initial feasible solutions up to 40% faster than SCDA. Figure 5.3 shows that LRCVaR is more scalable than Benders and SCDA to the increases in problem size. LRCVaR requires between 6% to 37% of the 60-minute time limit to find a feasible solution compared to Benders that requires between 16% to 100%, and SCDA between 8% to 60%.

The overtime probability threshold parameter, α_r, allows a decision-maker to adjust the emphasis between cost-effectiveness and variability of the optimal OR schedule [67]. Therefore, consistent performance of the proposed algorithm under different risk preferences is critical. Further numerical experiments are conducted to examine the quality of the proposed bounds for different α_r values. We solved all test instances in Table 5.2 and calculated the average optimality gap of the initial feasible solution. Table 5.5 shows that LRCVaR delivers a superior performance to the other methods for all α_r values. All algorithms find better feasible solutions in shorter run times as α_r increases. Figure 5.4 shows that the

Table 5.4: The quality of initial feasible solution using different algorithms

Instances	Benders		SCDA		LRCVaR	
	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
1	586.8	18.71	295.9	11.74	216	5.7
3	684	21.84	443.19	12.54	323.5	6.27
5	1080	23.06	538.68	14.14	402	6.9
7	1224	38.65	738.22	14.87	531.1	7.59
9	2304	36.56	1120.03	17.03	777.8	8.35
11	3240	33.36	1414.88	17.71	996.4	9.18
13	3486.9	30.42	1664.94	19.39	1197.8	10.1
15	3600	41.71	1879.2	20.79	1296	10.61
17	-	-	1953.6	21.92	1320	11.13
19	-	-	2144	23.96	1350	11.69

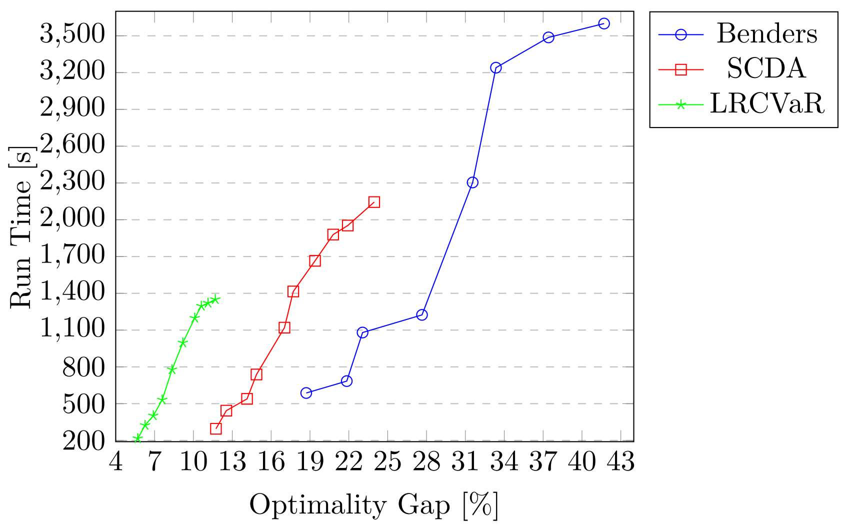


Figure 5.3: Comparing Benders, SCDA and LRCVaR in finding the initial feasible solution

quality of feasible solutions obtained by LRCVaR improves at a faster rate than Benders and SCDA.

Proposition 8. *CCP, CVaR and EV models provide the same optimal solution when* $\alpha_r = 1$.

Proof. When $\alpha_r = 1$, the chance constraints (5.2.15)-(5.2.16) can be written as

$$Pr\{tr_{krs} + \sum_i \delta_{is} y_{ikr} \leq cap_r : \forall k, s\} \geq 0 \quad \forall r \in R \tag{5.4.1}$$

Table 5.5: Quality of the initial feasible solution under different α_r values

α	Benders		SCDA		LRCVaR	
	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
10	2025.63	30.54	1183.22	17.70	841.06	8.75
20	1985.11	29.02	1124.06	16.70	773.78	7.09
30	1885.86	27.80	1067.86	15.86	704.14	5.21
40	1791.56	29.02	982.43	15.07	633.72	4.01
50	1719.9	26.57	913.66	14.32	576.69	3.03
60	1668.30	27.80	858.84	13.46	530.55	2.33
70	1634.94	28.41	807.31	12.84	482.8	1.55
80	1602.24	26.88	750.8	12.65	439.35	1.06
90	1522.13	26.31	705.75	11.88	404.2	0.64

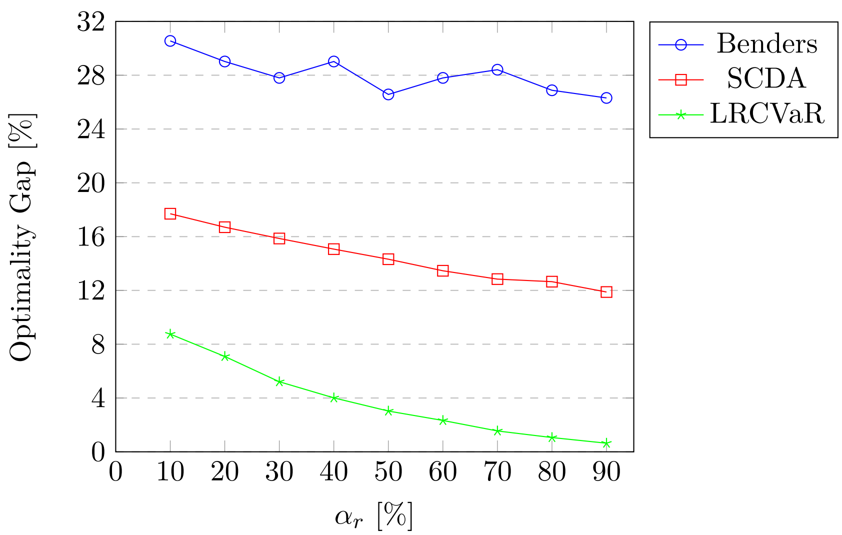


Figure 5.4: Quality of the initial feasible solution under different α_r values

which holds for all feasible solutions to the first-stage problem. Therefore, the chance constraints are redundant and we have M_{DEF} as a two-stage stochastic problem with relatively complete recourse. For $\alpha_r = 1$, VaR of second-stage costs is formulated as

$$VaR_1 = Min\left\{g(x,s) : CDF\left(g(x,s)\right) \geq 0\right\}. \tag{5.4.2}$$

Given that $CDF \geq 0$ for every random variable, we conclude

$$g(x,s) \geq VaR_1 \quad \forall x \in X, s \in \mathbb{S}. \tag{5.4.3}$$

Therefore, from Assumption 2 and inequality (5.4.3), the CVaR model minimizes the total costs over all scenarios that is equivalent to M_{DEF} without the chance constraints. We also know that the EV model is equivalent to the model M_{DEF} after relaxing the chance constraints. □

The previous observations of LRCVaR outperforming Benders and SCDA in solution quality and efficiency can be explained by Proposition 8. As α_r increases, the proposed lower and upper bounds in LRCVaR result in tighter approximations of model M_{DEF}. Moreover, the iterative process starts with a better initial feasible solution that is given by the CVaR approximation of model M_{DEF}. On the other hand, the performance improvements in Benders and SCDA is mainly attributed to the reduced number of generated feasibility cuts as the chance constraints are easier to satisfy when α_r increases.

Table 5.6 provides more details on the improved efficiency of LRCVaR as α_r approaches 1. LRCVaR Algorithm solves all test instances to optimality within one hour when $\alpha_r \geq$ 40%. Moreover, we solved all instances in Table 5.2 using Benders and SCDA for $\alpha_r = 20\%$, 40%, 60% and 80% and calculated the average solution time and optimality gap. Table 5.7 shows that SCDA and LRCVaR provide acceptable solutions within the 60-minute time limit. Figure 5.5 verifies the observations in Table 5.5 and Figure 5.4 stating that all algorithms provide better solutions in shorter times as α_r increases with LRCVaR showcasing a sharper decrease in solution time.

5.5 Conclusions

In this chapter, a two-stage chance-constrained mixed-integer programming model is proposed for the OR scheduling problem with stochastic surgery durations. The individual chance constraints control the risk of OR overtime. The goal is to minimize the sum of OR opening, OR overtime and patient waiting costs. We developed an augmented decomposition algorithm, LRCVaR, using strong lower and upper bounds. The bounds are obtained from Lagrangian relaxation and CVaR approximations of the proposed two-stage stochastic

Table 5.6: Performance of LRCVaR under different α_r values

Instances	$\alpha_r = 20\%$		$\alpha_r = 40\%$		$\alpha_r = 60\%$		$\alpha_r = 80\%$	
	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
1	389	0	251	0	181	0	130	0
2	534	0	326	0	225	0	191	0
3	580	0	384	0	330	0	229	0
4	667	0	437	0	348	0	230	0
5	830	0	498	0	402	0	262	0
6	865	0	573	0	463	0	302	0
7	1110	0	676	0	579	0	422	0
8	1310	0	826	0	598	0	498	0
9	1685	0	1074	0	852	0	630	0
10	2058	0	1191	0	866	0	715	0
11	2366	0	1445	0	1221	0	772	0
12	2522	0	1590	0	1179	0	850	0
13	2679	0	1891	0	1355	0	977	0
14	3160	0	2048	0	1736	0	1042	0
15	3590	0.41	2292	0	1910	0	1184	0
16	3600	0.88	2437	0	1764	0	1302	0
17	3600	1.20	2634	0	2264	0	1617	0
18	3600	1.72	2898	0	2084	0	1728	0
19	3600	1.75	3187	0	2293	0	1901	0
20	3600	2.24	3567	0	2645	0	2030	0

Table 5.7: Comparing the performance of Benders, SCDA and LRCVaR under different α_r values

α	Benders		SCDA		LRCVaR	
	Time (s)	Gap (%)	Time (s)	Gap (%)	Time (s)	Gap (%)
20	3471.98	17.7	3001.7	3.64	2117.25	0.41
40	3358.4	16.66	2770.75	3.18	1511.28	0
60	3182.14	15.63	2447.92	2.75	1164.8	0
80	2993.71	14.7	2126.77	2.38	850.62	0

model. Numerical experiments demonstrate that LRCVaR outperforms the Benders decomposition method and another decomposition algorithm, SCDA, in solving large-scale test instances. The proposed algorithm can solve instances with more than 200 surgeries and 40 ORs within 35 minutes. It is also shown that LRCVaR showcases a consistent behavior in providing high-quality solutions using different values for the overtime probability threshold parameter (α_r). The proposed bounds in LRCVaR enhance the performance of SCDA and achieve initial feasible solutions with 50% lower optimality gaps in shorter times. Moreover,

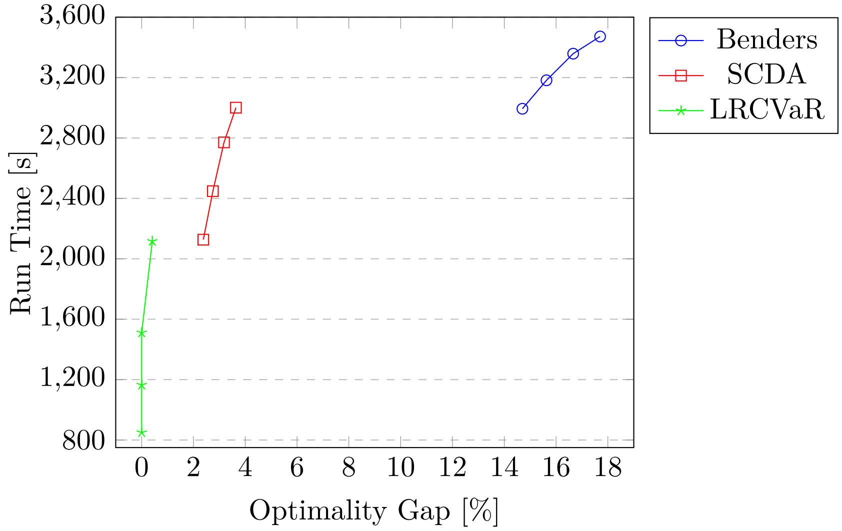


Figure 5.5: Comparing Benders, SCDA and LRCVaR in finding the initial feasible solution

LRCVaR achieved the initial feasible solutions up to 40% faster than SCDA. The numerical experiments were conducted on test instances where surgery durations follow log-normal distrubution functions. This can be a limitation of the current research. Therefore, evaluating the performance of LRCVaR on a variety of distribution functions can be an interesting topic for future research. Moreover, comparing different approximation methods in providing lower and upper bounds for the two-stage stochastic model can be another candidate topic for future research.

Chapter 6

Future Work

This research is directed toward addressing surgery duration uncertainty and providing risk-based stochastic models and solution methods that

- Maintain a trade-off between optimizing average cost and reducing variability of costs.
- Control the risk of undesirable events such as operating room overtime.
- Solve large-scale instances of the OR scheduling problem to optimality within reasonable time frames.

The proposed work can be enhanced in a variety of aspects. The remainder of this chapter identifies the limitations of the current work and proposes topics for future research.

6.1 Distributionally Robust Chance-Constrained Models for the Stochastic OR Scheduling Problem with Limited Upstream and Downstream Resources

In this research, we assumed that the surgery duration follows a known probability distribution. However, fitting proper distribution functions to the historical data, if available, may not be achieved for all cases [57]. It is also demonstrated that finding tailored cuts and bounds can accelerate the convergence to optimality significantly. Therefore, proposing a distributionally robust chance-constrained model coupled with efficient solution algorithms can ensure timely generation of OR schedules regardless of the underlying probability distribution of the random data.

6.2 Stochastic Models for the OR Scheduling Problem with Uncertain Surgery Duration and Emergency Patient Arrival

This research limits the scope of the proposed models to elective surgeries. It is also assumed that the upstream and downstream resources (e.g., nurses, ICU beds) are sufficiently available. It is shown that the random arrival of emergency patients to hospitals is one of the main sources of uncertainty in operating suites [79]. Moreover, it is shown that the majority of published articles have incorporated the availability of upstream and/or downstream resources in their work. Challenging the efficient decomposition algorithms proposed in this research by considering limited upstream and downstream resources as well as random patient arrivals can be an interesting topic for future research. The obtained solutions from this model can enable decision-makers to mitigate potential resource conflicts and implement integrated OR schedules.

References

[1] Ahmed, S. and Shapiro, A. (2008). Solving chance-constrained stochastic programs via sampling and integer programming. *Tutorials in Operations Research.* **10**:261–269.

[2] Artzner, P. and Delbaen, F. and Eber, J. M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance.* **9**(3):203–228.

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

[4] Batun, S. (2011). Scheduling multiple operating rooms under uncertainty (Doctoral dissertation, University of Pittsburgh).

[5] Batun, S. and Denton, B. T. and Huschka, T. R. and Schaefer, A. J. (2011). Operating room pooling and parallel surgery processing under uncertainty. *INFORMS Journal on Computing.* **23**(2):220–237.

[6] Begen, M. A. and Queyranne, M. (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research.* **36**(2):240–257.

[7] Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research.* **23**(4):769–805.

[8] Birge, J. R. and Louveaux, F. (2011). Introduction to stochastic programming. *Springer Science & Business Media.*

[9] Blake, J.T. and Carter, M.W. (1996). Surgical process scheduling: A structured review. *Journal of the Society for Health Systems.* **5**(3):17–30.

[10] Bruni, M. E. and Beraldi, P. and Conforti, D. (2015). A stochastic programming approach for operating theatre scheduling under uncertainty. *IMA Journal of Management Mathematics.* **26**(1):99–119.

[11] Cardoen, B. and Demeulemeester, E. and Beliën, J. (2006). Optimizing a multiple objective surgical case scheduling problem. *DTEW-KBI_0625.* 1–38.

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

[13] Cardoen, B. and Demeulemeester, E. and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research.* **201**(3):921–932.

[14] Cardoen, B. and Demeulemeester, E. and Van der Hoeven, J. (2010). On the use of planning models in the operating theatre: Results of a survey in Flanders. *International Journal of Health Planning and Management.* **25**(4):400–414.

[15] Charnes, A. and Cooper, W. W. (1959). Chance-constrained programming. *Management science.* **6**(1):73–79.

[16] Chen, X. and Shapiro, A. and Sun, H. (2019). Convergence analysis of sample average approximation of two-stage stochastic generalized equations. *SIAM Journal on Optimization.* **29**(1): 135–161.

[17] Cima, R. and Brown, M. and Hebl, J. and Moore, R. and Rogers, J. and Kollengode, A. and Amstutz, G. and Weisbrod, C. and Narr, B. and Deschamps, C. and Team, S.P.I (2011). Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *Journal of the American College of Surgeons.* **213**(1):83–92.

[18] Deng, Y. and Shen, S. and Denton, B. (2014). Chance-constrained surgery planning under uncertain or ambiguous surgery duration. *Available at SSRN, 2432375.*

[19] Deng, Y. and Shen, S. (2016). Decomposition algorithms for optimizing multi-server appointment scheduling with chance constraints. *Mathematical Programming.* **157**(1):245–276.

[20] Deng, Y. and Shen, S. and Denton, B. (2016). Chance-constrained surgery planning under uncertain or ambiguous surgery duration. *Available at SSRN, 2432375.*

[21] Deng, Y., Shen, S., and Denton, B. (2019). Chance-constrained surgery planning under conditions of limited and ambiguous data. *INFORMS Journal on Computing.* **31**(3):559–575.

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

[23] Denton, B. and Viapiano, J. and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science.* **10**(1):13–24.

[24] Denton, B. T. and Miller, A. J. and Balasubramanian, H. J. and Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research.* **58**(4-part-1):802–816.

[25] Fei, H. and Meskens, N. and Chu, C. (2006). An operating theatre planning and scheduling problem in the case of a block scheduling strategy. *IEEE 2006 International Conference on Service Systems and Service Management.* **1**:422–428.

[26] Fisher, M. L. (1973). Optimal solution of scheduling problems using Lagrange multipliers: Part I. *Operations Research.* **21**(5):1114–1127.

[27] Fisher, M. L. (2004). The Lagrangian relaxation method for solving integer programming problems. *Management Science.* **50**:1861–1871.

[28] Fisher, M. L. and Shapiro, J. F. (1974). Constructive duality in integer programming. *SIAM Journal on Applied Mathematics.* **27**(1):31–52.

[29] Freeman, N. K. and Melouk, S. H. and Mittenthal, J. (2015). A Scenario-Based approach for operating theater scheduling under uncertainty. *Manufacturing & Service Operations Management.* **18**(2):245–261.

[30] Gerchak, Y. and Gupta, D. and Henig, M. (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science.* **42**(3):321–334.

[31] Girotto, J. A. and Koltz, P. F. and Drugas, G. (2010). Optimizing your operating room: or, why large, traditional hospitals don't work. *International Journal of Surgery.* **8**(5):359–367.

[32] Gocgun, Y. and Ghate, A. (2012). Lagrangian relaxation and constraint generation for allocation and advanced scheduling. *Computers & Operations Research.* **39**(10):2323–2336.

[33] Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science.* **14**(1):89–114.

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

[35] Gul, S. and Denton, B. T. and Fowler, J. W. and Huschka, T. (2011). BiâĂŘCriteria scheduling of surgical services for an outpatient procedure center. *Production and Operations management.* **20**(3):406–417.

[36] M'Hallah, R. and Visintin, F. (2019). A stochastic model for scheduling elective surgeries in a cyclic Master Surgical Schedule. *Computers & Industrial Engineering.* **129**:156–168.

[37] Hampel, F. R. and Ronchetti, E. M. and Rousseeuw, P. J. and Stahel, W. A. (2011). Robust statistics: the approach based on influence functions. *John Wiley & Sons.* **196**

[38] Herring, W. L. and Herrmann, J. W. (2011). A stochastic dynamic program for the single-day surgery scheduling problem. *IIE Transactions on Healthcare Systems Engineering.* **1**(4):213–225.

[39] Jebali, A. and Alouane, A. B. H. and Ladet, P. (2006). Operating rooms scheduling. *International Journal of Production Economics.* **99**(1):52–62.

[40] Jebali, A. and Diabat, A. (2017). A Chance-constrained operating room planning with elective and emergency cases under downstream capacity constraints. *Computers & Industrial Engineering.* **114**:329–344.

[41] Kamran, M. A. and Karimi, B. and Dellaert, N. (2018). Uncertainty in advance scheduling problem in operating room planning. *Computers & Industrial Engineering.* **126**:252–268.

[42] Khabazian, A. and Zaghian, M. and Lim, G. J. (2019). A feasibility study of a risk-based stochastic optimization approach for radiation treatment planning under setup uncertainty. *Computers & Industrial Engineering.* **135**:67–78.

[43] Kisiala, J. (2015). Conditional value-at-risk: Theory and applications. *arXiv preprint arXiv:1511.00140.*

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

[45] Koeleman, P. M. and Koole, G. M. (2012). Optimal outpatient appointment scheduling with emergency arrivals and general service times. *IIE Transactions on Healthcare Systems Engineering.* **2**(1):14–30.

[46] Küçükyavuz, S. and Sen, S. (2017). An introduction to two-stage stochastic mixed-integer programming. *INFORMS Tutorials in Operations Research.* 1–27.

[47] Latorre-Núñez, G. and Lüer-Villagra, A. and Marianov, V. and Obreque, C. and Ramis, F. and Neriz, L. (2016). Scheduling operating rooms with consideration of all resources, post anesthesia beds and emergency surgeries. *Computers & Industrial Engineering.* **97**:248–257.

[48] Lamiri, M. and Xie, X. and Dolgui, A. and Grimaud, F. (2008). A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research.* **185**(3):1026–1037.

[49] Lamiri, M. and Grimaud, F. and Xie, X. (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics.* **120**(2):400–410.

[50] Leeftink, G. and Hans, E. W. (2018). Case mix classification and a benchmark set for surgery scheduling. *Journal of scheduling.* **21**(1):17–33.

[51] Lim, G. J. and Mobasher, A. and Bard, J. F. and Najjarbashi, A. (2016). Nurse scheduling with lunch break assignments in operating suites. *Operations Research for Health Care.* **10**:35–48.

[52] Lim, G. J. and Kardar, L. and Ebrahimi, S. and Cao, W. (2020). A risk-based modeling approach for radiation therapy treatment planning under tumor shrinkage uncertainty. *European Journal of Operational Research.* **280**(1):266–278.

[53] Luedtke, J. (2014). A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming.* **146**:219–244.

[54] Liu, X. and Küçükyavuz, S. and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming.* **157**(1):219–243.

[55] Macario, A. and Dexter, F. and Traub, R. D. (2001). Hospital profitability per hour of operating room time can vary among surgeons. *Anesthesia & Analgesia.* **93**(3):669–675.

[56] Macario, A. (2006). Are your operating rooms efficient? *Anesthesiology.* **105**(2):237–240.

[57] Macario, A. (2006). What does one minute of operating room time cost? *Journal of Clinical Anesthesia.* **22**(4):233–236.

[58] Magerlein, J.M. and Martin, J.B. (1978). Surgical demand scheduling: a review. *Health Services Research.* **13**(4):418–433.

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

[60] Markowitz, H. (1952). Portfolio selection. *The journal of finance.* **7**(1):77–91.

[61] May, J.H. and Spangler, W.E. and Strum, D.P. and Vargas, L.G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management.* **20**(3):392–405.

[62] McManus, M. L. and Long, M. C. and Cooper, A. and Mandell, J. and Berwick, D. M. and Pagano, M. and Litvak, E. (2003). Variability in surgical caseload and access to intensive care services. *Anesthesiology: The Journal of the American Society of Anesthesiologists.* **98**(6):1491–1496.

[63] Min, D. and Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research.* **206**(3):642–652.

[64] Molina-Pariente, J. M. and Hans, E. W. and Framinan, J. M. (2016). A stochastic approach for solving the operating room scheduling problem. *Flexible Services and Manufacturing.* 1–28.

[65] Molina-Pariente, J. M. and Hans, E. W. and Framinan, J. M. (2018). A stochastic approach for solving the operating room scheduling problem. *Flexible services and manufacturing journal.* **30**(1-2):224–251.

[66] Muñoz, E. and Muñoz III, W. and Wise, L. (2010). National and surgical health care expenditures, 2005-2025. *Annals of surgery.* **251**(2):195–200.

[67] Najjarbashi, A. and Lim, G. J. (2019). A variability reduction method for the operating room scheduling problem under uncertainty using CVaR. *Operations Research for Health Care.* **20**:25–32.

[68] Najjarbashi, A. and Lim, G. J. (2020)., A Decomposition Algorithm for the Two-Stage Chance-Constrained Operating Room Scheduling Problem. *in IEEE Access.* **8**:80160–80172.

[69] Nemirovski, A. and Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization.* **17**(4):969–996.

[70] Neyshabouri, S. and Berg, B. P. (2017). Two-stage robust optimization approach to elective surgery and downstream capacity planning. *European Journal of Operational Research.* **260**(1):21–40.

[71] Noorizadegan, M. and Seifi, A. (2018). An efficient computational method for large scale surgery scheduling problems with chance constraints. *Computational Optimization and Applications.* **69**(2):535–561.

[72] Noyan, N. (2012). Risk-averse two-stage stochastic programming with an application to disaster management. *Computers & Operations Research.* **39**(3):541–559.

[73] Pang, B. and Xie, X. and Song, Y. and Luo, L. (2018). Surgery Scheduling Under Case Cancellation and Surgery Duration Uncertainty. *IEEE Transactions on Automation Science and Engineering.* **16**(1):74–86.

[74] Pratt, J.W. (1992). Risk aversion in the small and in the large. *Foundations of Insurance Economics.* **13**:83–98.

[75] Przasnyski, Z.H. (1986). Operating room scheduling: A literature review. *AORN Journal.* **44**(1):67–82.

[76] Rahmaniani, R. and Crainic, T. G. and Gendreau, M. and Rei, W. (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research.* **259**(3):801–817.

[77] Rath, S. and Rajaram, K. and Mahajan, A. (2017). Integrated anesthesiologist and room scheduling for surgeries: methodology and application. *Operations Research.* **65**(6):1460–1478.

[78] Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk.* **2**:21–42.

[79] Samudra, M. and Van Riet, C. and Demeulemeester, E. and Cardoen, B. and Vansteenkiste, N. and Rademakers, F. E. (2016). Scheduling operating rooms: Achievements, challenges and pitfalls. *Journal of Scheduling.* **19**(5):493–525.

[80] Sarin, S. C. and Sherali, H. D. and Liao, L. (2014). Minimizing conditional-value-at-risk for stochastic scheduling problems. *Journal of Scheduling.* **17**(1):5–15.

[81] Schmid, V. and Doerner, K. F. (2013). Examination and operating room scheduling including optimization of intrahospital routing. *Transportation Science.* **48**(1):59–77.

[82] Shapiro, J. F. (1971). Generalized Lagrange multipliers in integer programming. *Operations Research.* **19**(1):68–76.

[83] Shapiro, A. and Dentcheva, D. and Ruszczyński, A. (2009). Lectures on stochastic programming: modeling and theory. *Society for Industrial and Applied Mathematics.*

[84] Shylo, O. V. and Prokopyev, O. A. and Schaefer, A. J. (2012). Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing.* **25**(4):682–692.

[85] Smith, C. and Spackman, T. and Brommer, K. and Stewart, M. and Vizzini, M. and Frye, J. and Rupp, W. (2013). Re-engineering the operating room using variability methodology to improve health care value. *Journal of the American College of Surgeons.* **216**(4):559–568.

[86] Smith-Daniels, V.L. and Schweikhart, S.B. and Smith-Daniels, D.E. (1988). Capacity management in health care services: Review and future research directions. *Decision Sciences.* **19**(4):889–919.

[87] Strum, D. P. and May, J. H. and Vargas, L. G. (2000). Modeling the uncertainty of surgical procedure times–comparison of log-normal and normal Models. *The Journal of the American Society of Anesthesiologists.* **92**(4):1160–1167.

[88] Uryasev, S. (2000). Conditional value-at-risk: Optimization algorithms and applications. *In Computational Intelligence for Financial Engineering, 2000.(CIFEr) Proceedings of the IEEE/IAFE/INFORMS 2000 Conference.* 49–57.

[89] Van Huele, C. and Vanhoucke, M. (2013). Analysis of the integration of the physician rostering problem and the surgery scheduling problem. *Journal of medical systems.* **38**(6):43.

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

[91] Wang, Y. and Tang, J. and Fung, R. Y. (2014).A column-generation-based heuristic algorithm for solving operating theater planning problem under stochastic demand and surgery cancellation risk. *IEEE International Journal of Production Economics.* **158**: 28–36.

[92] Wang, S. and Li, J. and Peng, C. (2017). Distributionally robust chance-constrained program surgery planning with downstream resource. *IEEE International Conference In Service Systems and Service Management (ICSSSM).* 1–6.

[93] Wang, Y. and Zhang, G. and Zhang, L. and Tang, J. and Mu, H. (2018). A Column-Generation Based Approach for Integrating Surgeon and Surgery Scheduling. *IEEE Access.* **6** :41578–41589.

[94] Weiss, A. J. and Elixhauser, A. and Andrews, R. M. (2014). Characteristics of operating room procedures in US hospitals, 2011. *Agency for Healthcare Research and Quality (US), Rockville, MD.* statistical brief #170.

[95] Xiang, W. and Yin, J. and Lim, G. (2015). A short-term operating room surgery scheduling problem integrating multiple nurses roster constraints. *Artificial Intelligence in Medicine.* **63**(2):91–106.

[96] Xiao, G. and Van Jaarsveld, W. and Dong, M. and Van De Klundert, J. (2016). Stochastic programming analysis and solutions to schedule overcrowded operating rooms in China. *Computers & Operations Research.* **74**:78–91.

[97] Zaghian, M. and Lim, G. J. and Khabazian, A. (2018). A chance-constrained programming framework to handle uncertainties in radiation therapy treatment planning. *European Journal of Operational Research.* **266**(2):736–745.

[98] Zhang, Z. and Denton, B. and Xie, X. (2015). Branch and price for chance constrained bin packing. *Available at Optimization-Online* $http://www.optimization-online.org/DB\_HTML/2015/11/5217.html$.

[99] Zhou, B. H. and Yin, M. and Lu, Z. Q. (2016). An improved Lagrangian relaxation heuristic for the scheduling problem of operating theatres. *Computers & Industrial Engineering.* **101**:490–503.

Appendix A

Robust Measures of Variability for Numerical Experiments in Chapter 3

A.1 Interquartile Range:

The IQR measure can be calculated as

$$IQR = Q_3 - Q_1 = CDF^{-1}(0.75) - CDF^{-1}(0.25)$$

where Q_1 and Q_3 are the first and the third quartile of the cost vector, respectively.

A.2 Median Absolute Deviation:

The MAD measure can be calculated as

$$MAD = median(|\chi_s - median(XX)|)$$

where χ_s and XX are the total cost in scenario s and the cost vector, respectively.