Robust Multi-Class Multi-Period Scheduling of MRI Services with Wait Time Targets

by

Akram Mirahmadi Shalamzari

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Sciences
in
Management Sciences

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In recent years, long wait times for healthcare services have become a challenge in most healthcare delivery systems in Canada. This issue becomes even more important when there are priorities in patients' treatment which means some of the patients need emergency treatment, while others can wait longer. One example of excessively long wait times in Canada is the MRI scans. These wait times are partially due to limited capacity and increased demand, but also due to sub-optimal scheduling policies. Patients are typically prioritized by the referring physician based on their health condition, and there is a wait time target for each priority level. The difficulty of scheduling increases due to uncertainty in patients' arrivals and service times. In this thesis, we develop a multi-priority robust optimization (RO) method to schedule patients for MRI services over a multi-period finite horizon. First, we present a deterministic mixed integer programming model which considers patient priorities, MRI capacity, and wait time targets for each priority level. We then investigate robust counterparts of the model by considering uncertainty in patients' arrivals and employing the notion of the budget of uncertainty. Finally, we apply the proposed robust model to a set of numerical examples and compare the results with those of the non-robust method. Moreover, sensitivity analysis is performed over capacity, penalty cost, service level, and budget of uncertainty. Our results demonstrate that the proposed robust approach provides solutions with higher service levels for each priority, and lower patients' wait time in realistic problem instances. The analysis also provides some insights on expanding capacity and choosing the budget of uncertainty as a trade-off between performance and conservatism.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Houra Mahmoudzadeh and Hossein Abouee-Mehrizi for their continuous guidance and support throughout my study and research.

I also would like to thank my reading committee, Professor Jim Bookbinder, and Professor Qi-Ming He for their insightful comments and suggestions.

A very special thanks to my friends Gohram and Milad for their valuable help and support throughout my study.

Finally, a warm thanks to my dearest friends Nazanin, Keyvan, Parisa, Saeede, Leila, Danial, Niloofar, and Majid who have always supported and helped me in many different ways.

Dedication

This work is dedicated to my beloved mom and dad, my sister Elahe and my brothers Mahdi and Hamid. All I have and will accomplish are only possible due to their love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A magnetic resonance imaging (MRI) scan is an important step in the early diagnosis of many medical conditions including the brain, neck, and spinal cord diseases. In some cases, MRI provides essential information that cannot be obtained through other medical imaging methods. Most of the people may need MRI scans at least once in their lifetime (CADTH, 2015).

Because of the growing and aging population, demand for MRI examinations is high and has been increasing in the past few years (CIHI, 2016). According to Canadian Institute for Health Information (CIHI), 1.95 million MRI examinations were performed in Canada in 2015. 40% of these patients experienced difficulty in getting specialized diagnostic tests, while this percentage is 21% internationally (CIHI, 2016). Based on the reported data, since 2012, patients' wait times for MRI exams have increased (Figure 1.1). The CIHI report shows that 90% of patients wait an average of 105 days to receive an MRI service (CIHI, 2016). These excessive wait times become even more vital in some urgent cases, where it may potentially impact the health outcomes. Often, there are different medically acceptable wait times for different types of patients. Some patients need urgent and immediate service while others can wait up to several weeks.

In Canada, patients who need MRI scans are categorized into four priorities based on assessments performed by their primary physicians (CAR, 2013). Ontario has determined optimal wait time targets within which a patient should be served. Patients with priority 1, the highest priority level, must be served in less than 1 day and patients with priority 2 should receive MRI in less than 2 days. Priorities 3 and 4 can wait up to 10 and 28 days, respectively (MOHLTC, 2017a). These wait time targets are developed by Canadian Association of Radiologists (CAR) in collaboration with the Ontario government (WTA, 2013). In 2004, the First Ministers of Canada committed to decreasing patient wait times as part of a 10-year plan to improve health care (CIHI, 2017). They set a mandated target to serve at least 90% of patients with different priorities within their wait time targets in a fiscal year (365 days) (CAR, 2013).

Based on the data from 74 MRI hospitals in Ontario, among all patients who need MRI, 2% are classified as priority 1, 7% priority 2, 14% priority 3, and 77% are priority 4 (lowest priority) (Jiang, 2016). Since waiting too long for MRI scan may have severe impacts on patients' health outcomes, managers and policy makers in health care systems try to reduce patients' wait times by improving appointment scheduling. However, these goals are not met in current practice. The percentage of patients who receive the service in their wait time targets are shown in Table 1.1, based on the report of Ontario ministry of health and long-term care (MOHLTC, 2017b). The primary motivation of our work is to provide a method to reduce wait times in an MRI section.

In a system with limited capacity, there is often a queue. There are three ways to improve the performance of these systems; (1) Increasing the capacity is the simplest way; however, it is costly. (2) Controlling patients' arrivals by announcing the wait times

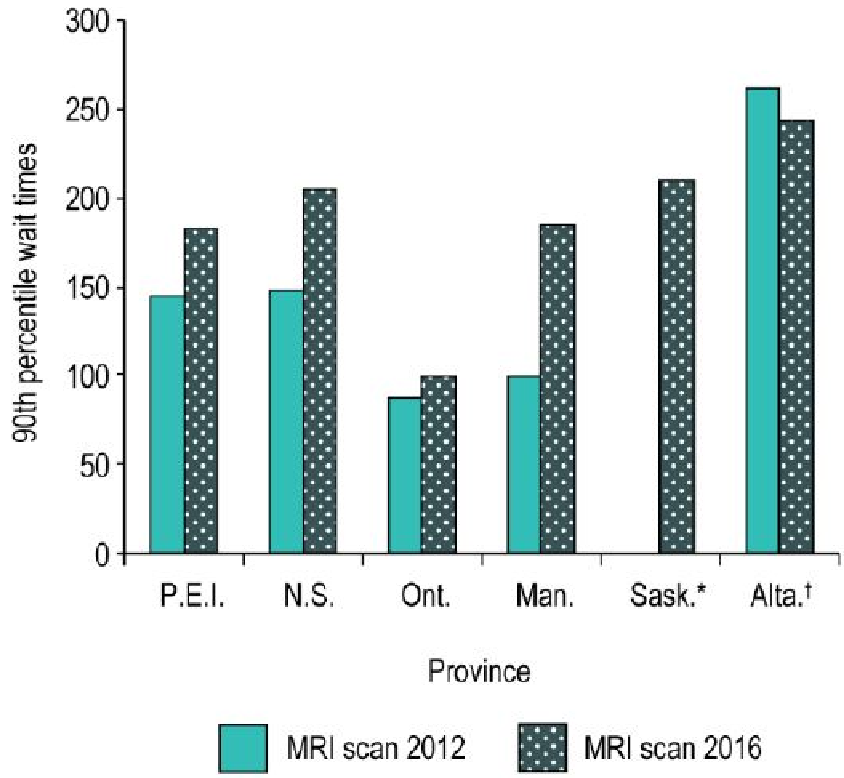


Figure 1.1: Provincial wait times (in days) for MRI exams, April to September 2012 and 2016 (CIHI, 2017)

∗ Saskatchewan did not report data in 2012.

+ Alberta excluded the data from Edmonton region in 2012. In 2016 the data of all regions are included.

through, for example, an online system (Dong et al., 2015). (3) Regulating the queue by efficient scheduling policies. The imbalance between available resources and demand usually causes capacity deficiencies that can be decreased by resource allocation and capacity optimization. An improved patient scheduling would create a more efficient capacity allocation and a reduction in wait times (Chakraborty et al., 2010). In this thesis, our focus is mainly on improving the performance through regulating the queue. While this thesis focuses on the scheduling of multi-priority patients in an MRI section where different priorities have different wait time targets, our approach may be applicable to a broader domain, such as

Provincial wait time by priority- Feb 2017			
MRI service	90th percentile wait time (days)	Wait time target (days)	% Completed within wait time target
Priority 1	N/A	1	N/A
Priority 2	4	2	81%
Priority 3	42	10	61%
Priority 4	105	28	39%

Table 1.1: Provincial wait time by priority- Feb 2017 (MOHLTC, 2017b)

scheduling patients for operating rooms (Kim and Horowitz, 2002), and generally, when there is a limited resource with multi-priority levels and related wait time targets.

There are two commonly-used appointment scheduling methods in healthcare centers; first-come, first-served and strict policy (Gupta and Denton, 2008). In first-come, first-served, there is a single line of patients, and all patients with different priorities are assumed to be equal regarding the urgency of service. In this method, the order of receiving service is based on patients' arrival times. This scheduling method is not practical for MRI because priorities and related wait times are neglected. On the other hand, in the strict policy, higher priorities would receive the service first, and as long as the queues of higher priority groups are not empty, lower priority groups would not receive the service. However, hospitals aim to serve lower priorities, as well, as long as higher priorities are being served within their wait time targets (Jiang, 2016).

In addition, due to the existing uncertainties in demands for MRI examination and duration of MRI services, the scheduling is more challenging. Two most popular approaches to deal with uncertainties in data and information are Stochastic Optimization (SO) and Robust Optimization (RO). In practice, it is hard to obtain the exact distribution of demand for MRI scans. SO needs the information of the distribution of the random variable, while RO does not necessarily require the distribution information and it assumes that the uncertain data or parameter belongs to an "uncertainty set" (Ben-Tal et al., 2009). Also, in contrast to stochastic optimization which finds solutions that are protected from some probabilistic sense in stochastic uncertainty, robust optimization seeks to build a solution that is feasible for any realization of uncertainty within a given uncertainty set (Bertsimas et al., 2011).

We aim to develop a scheduling method to decrease patients' wait times in the system. Our scheduling approach is an advance scheduling, which means the number of patients that are going to be served from each priority in each day is determined at time $t = 0$ of the horizon. Then, patients from each priority are served by first-come, first-serve policy within the same day. The goal is to minimize the sum of penalty costs of delaying patients while serving at least 90% of patients within their wait time targets. Then, we apply robust optimization methods to incorporate the uncertainties in demand for MRI scans into our proposed scheduling method.

In this thesis, we develop robust counterparts of a capacitated multi-period scheduling problem with multi-priority patients subject to arrival uncertainty. The rest of the thesis is as follows; Section 2 presents a literature review on healthcare scheduling and robust optimization. In Section 3, the problem is defined, and a mixed integer program is proposed to schedule multi-priority patients over a multi-period finite horizon. Then, uncertainty in the patient's arrival is considered, and robust counterparts of the proposed model are presented. In Section 5, we present numerical results and sensitivity analysis over important parameters of the model. Finally, in Section 6, concluding remarks and future research directions are presented.

Chapter 2

Related Work

2.1. Healthcare Scheduling

In any system with a limited capacity, when there is a mismatch between capacity and demand, there is a queue. There are three ways to improve the performance of these systems; First, by increasing the capacity, which is costly. Second, by controlling the arrivals of the system with online wait time announcements (Dong et al., 2015). Third, by regulating the queue and implementing a good appointment scheduling policy. Chakraborty et al. (2010) state that inadequate scheduling leads to inefficiency in operational management, whereas a good scheduling policy would reduce the wait time and positively optimize the access to the service, also, good scheduling systems that control the flow of patients are the primary means to improve the efficiency of resource utilization (Lowery and Martin, 1989). There is plenty of research on appointment scheduling. For comprehensive reviews, readers are encouraged to refer to Cayirli and Veral (2003), Denton and Gupta (2003), Mondschein and Weintraub (2003), Cardoen et al. (2010) and Guerriero and Guido (2011). However, the problem of allocating the medical capacity when there are multiple priorities or classes has received limited attention. In general, scheduling systems can be divided into two areas; Advance scheduling and allocation scheduling (Magerlein and Martin, 1978). Advance scheduling refers to when patients are scheduled in the future in advance of the service date, and the latter refers to the methods that schedule patients on the day of appointment based on a waiting list. Our work falls into the first category and mainly focuses on multi-period advance scheduling for multi-priority patients with different wait time targets.

Allocation scheduling has received a lot of attention in recent years. Gerchak et al. (1996) define a stochastic dynamic program to schedule elective surgeries on each day when there are two classes of patients to utilize the capacity; emergency and elective surgeries. They incorporate the uncertain duration of surgery to in their analysis and determine the number of elective surgeries on each day. Green et al. (2006) investigate the problem of scheduling patients for diagnostic services. They develop a method to schedule patients within a day, after determining a fixed number of outpatient exams.

Several heuristic rules are developed to determine which patients to serve next when there are patients who are waiting in both inpatient and outpatients categories. Qu et al. (2007) develop a scheduling method to find the percentage of open-access appointments that are required to satisfy the patients' demand in each session. They also evaluate the sensitivity of solutions under changes of available capacity, demand distribution and no-show rate. Gupta and Wang (2008) propose a Markov decision process model to schedule patients in a primary-care clinic. They consider patients' preference for a specific time of day and a specific physician. In their approach, allocated spots are dynamically changed to meet the demands. They develop several heuristics to maximize the revenue by deciding which appointment request to accept. Ayvaz and Huh (2010) also investigate the problem of resource allocation in a hospital. They design a model by adopting a dynamic programming approach and propose a heuristic to solve the problem with multiple types of patients that have different reactions to the delays in service.

In Min and Yih (2010), the problem of scheduling different priorities of patients for elective surgeries is investigated. They incorporate the uncertainties in surgery duration and the capacity of the intensive care unit. The authors develop a stochastic dynamic program to solve the problem. Luo et al. (2012) study the problem of scheduling patients' appointments under no-show behavior. They develop a heuristic by considering policies where a scheduled patient can be interrupted by an emergency request. Feldman et al. (2014) also develop heuristics for scheduling appointments, assuming that patients' preferences of a particular time for appointment follow a multinomial logit model. Patients can choose among offered days or wait for a later appointment date. They also incorporate the case of no-show behavior in their approach.

Priority queue is one of the methods in dynamic appointment scheduling, in which the order of patients' treatment is determined by their priorities. Kleinrock and Finkelstein (1967) present a queuing policy as "time-dependent priority queue" to deal with the drawback of classical priority queue discipline. Their proposed discipline moderates the issue of not serving lower priorities before higher priorities. In their policy, a lower priority patient can reach an accumulated priority to receive the service before a higher one. This accumulated priority is a linear function of patients' wait time in the system. Stanford et al. (2014) improve the approach proposed by Kleinrock and Finkelstein (1967) by considering multiple classes of patients with performance targets. They derive the wait time distributions of patients in different classes to obtain the "maximum priority process". Li (2015) extends the theory of accumulating priority queue (APQ) in Kleinrock and Finkelstein (1967) by considering a multi-class, multi-server queue. The author proposes a discipline to minimize the weighted average of the expected excess wait time of patients.

Compared to allocation scheduling, research focused on advance scheduling is limited. It is mainly because of the computational effort needed in advance scheduling to consider the capacity restriction over slots in the booking horizon, as opposed to allocation scheduling which is required to focus on the current time slot (Gocgun and Puterman, 2014).

Patrick et al. (2008) study a dynamic scheduling method for a diagnostic facility, which consists of two types of patients' demand: inpatients and outpatients. They formulate the model as a Markov decision process (MDP). Since the size of the state space makes MDP a non-tractable approach, they employ approximate dynamic programming to solve the model. They also use simulation to evaluate the solution quality. However, Patrick et al. (2008) do not employ the idea of wait time targets explicitly. Their method tries to serve patients in a vital class before the particular target date. Gocgun and Ghate (2012) extend the model of Patrick et al. (2008) by considering multiple resources and then formulate the allocation and advance scheduling problems as Markov decision process models. They propose a heuristic based on approximate dynamic programming which employs Lagrangian relaxation and constraint generation to obtain the schedule.

Saure et al. (2012) consider the problem of patient scheduling for radiation therapy by implementing the approach of Patrick et al. (2008). Gocgun and Puterman (2014) extend the work of Saure et al. (2012) to an application in chemotherapy appointment scheduling by considering target days and tolerance around those targets. The proposed tolerance provides a degree of flexibility for appointment booking of patients. Truong (2015) study analytical properties of a dynamic advance scheduling problem of two classes of patients; emergency and regular with multiple resources. The author also develops an efficient algorithm to calculate the optimal policy. Astaraky and Patrick (2015) develop an advance scheduling method for the problem of multi-class patient scheduling for surgeries with multiple resources. Their proposed schedule aims to minimize the sum of the interval between patient's request and date of surgery, operating room's overtime, and congestion in the wards. To find the schedule, a simulation-based approximate dynamic programming method is presented. Jiang (2016) proposes two dynamic scheduling policies: weight accumulation and priority promotion, for scheduling patients for MRI scans. The author shows a substantial reduction in patients' wait times in the system. Parizi and Ghate (2016) study a class of advance scheduling problems with multiple classes of patients that need multiple resources. They also consider no-show behavior, cancellation, and overbooking in their studies. They develop a Markov decision process model and employ approximate dynamic programming to find a schedule.

From an advance scheduling perspective, there have been some studies addressing challenges associated with scheduling of multi-priority patients or considering patients with wait time targets. However, to the best of our knowledge, no study addresses multi-period scheduling of multi-priority patients which explicitly considers different wait time targets for each priority level. In the present study, we investigate the problem of scheduling multi-priorities patients with different wait time targets in an MRI section. Our scheduling approach is an advance scheduling. It determines the number of patients from each priority in each day that can receive service. The goal is to minimize the sum of penalty cost of delaying patients while serving at least 90% of patients within their wait time targets.

2.2. Robust Optimization

One of the most important things to consider when developing a model is the uncertainty in data and information. In the real world, we are facing problems in which determining the exact values of parameters is difficult and in some cases impossible. Therefore, to address these issues, it is necessary to exploit methods that can incorporate the uncertainties in data into the model and apply them to increase the reliability and accuracy of the results. Accordingly, various methods can be used to consider data uncertainty in a mathematical programming model. These include robust optimization (RO) and stochastic optimization (SO).

In some cases, despite the incompleteness of the information and lack of accuracy in data, robust optimization provides better results, while other methods sometimes face feasibility and tractability problems that make them not applicable in practice. For example, in stochastic optimization, the actual probability distribution of uncertain parameter is assumed to be known or estimated, and the goal is to optimize an expected measure (Birge and Louveaux, 2011). However, the lack of the availability of the distribution function of the parameters makes this problem difficult for the analyst to handle. On the other hand, RO doesn't necessarily need the distribution information, and it assumes that the uncertain data or parameter belongs to an "uncertainty set". An attractive feature of RO is that for many classes of uncertainty sets, it leads to a computationally tractable problem and the robust counterpart can be reduced to an equivalent mathematical program which is not much harder than the original (non-robust) problem to solve (Ben-Tal et al., 2009). Furthermore, RO does not suffer from the curse of dimensionality, which is the difficulty in some dynamic programming problems. Especially, many researchers consider robust optimization as a tractable approximation approach that can solve some stochastic programming models which are considered impossible to solve (Chen et al., 2007). For a detailed review of SO topics, readers are encouraged to see Birge and Louveaux (2011).

Robust optimization was introduced by El Ghaoui et al. (1998), Ben-Tal and Nemirovski (1998), and Bertsimas and Sim (2004). The idea of using a "maximin" model is not a new idea. Wald (1945) for the first time introduced this idea which is influenced by the notion of zero-sum games in game theory. Later on, Soyster (1973) was the first one who developed linear models that are protected against the uncertainty in data. He presents a linear optimization model to find the solution which is feasible for all data that belong in a convex set, but the problem is that the result of his model is too conservative and it gives up too much optimality. The primary model of Soyster is as follows:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \sum_{j=1}^{n} \tilde{\mathbf{A}}_j \mathbf{x}_j \leq \mathbf{b}, & \qquad \forall \tilde{\mathbf{A}}_j \in K_j,\ j = 1, 2, ..., n \\ \mathbf{x} \geq 0 & \end{aligned}$$

In this model, each column $\tilde{\mathbf{A}}_j$ of matrix $\tilde{\mathbf{A}}$ belongs to the uncertainty set K_j, which is convex. Soyster shows that the problem is equivalent to the following:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \sum_{j=1}^{n} \bar{\mathbf{A}}_j \mathbf{x}_j \leq \mathbf{b}, & \qquad j = 1, 2, ..., n \\ \mathbf{x} \geq 0 & \end{aligned}$$

In this model $\bar{a}_{ij} = \sup_{\tilde{\mathbf{A}}_j \in K_i} (\tilde{\mathbf{a}}_{ij})$, where $\bar{a}_{ij}$ are the entries of matrix $\bar{\mathbf{A}}$. In other words, Soyster considers the maximum value of matrix $\bar{\mathbf{A}}$ and puts that in the constraints. Therefore, the solution is too conservative since it is protected against the worst-case scenario.

In general, the issue of over-conservatism in robust optimization is important, as the worst-case solution is derived from an uncertain parameter set. Thus, choosing an appropriate uncertainty set to achieve a trade-off between the protection against uncertainty and performance is essential. In some cases, the worst-case is computed over a finite number of scenarios like historical data or over continuous, convex uncertainty sets, such as polyhedral or ellipsoidal uncertainty (Bertsimas et al., 2011).

Ben-Tal and Nemirovski (1998) reintroduce the concept of robust optimization for convex optimization problems. Their assumption of uncertainty was of that enclosed within an ellipsoid. Then, Ben-Tal and Nemirovski (1999) introduce RO for linear programming (LP) problem by assuming the uncertainty set as an ellipsoidal set. Ben-Tal and Nemirovski (2000) demonstrate that the robust counterpart (RC) of the LP with an ellipsoidal uncertainty set is computationally tractable, and the equivalent robust counterpart can be reformulated as a conic quadratic program, which has higher computational cost than solving an LP (Ben-Tal et al., 2009). Their nominal model is as follows:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned}$$

Where $\mathbf{A} = (a_{ij})$ denotes the uncertain parameters of the model. They present the following robust problem:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \sum_j a_{ij}x_j + \sum_{j \in J_i} \hat{a}_{ij}y_j + \Omega_i \sqrt{\sum_{j \in J_i} \hat{a}_{ij}^2 z_{ij}^2} \leq b_i && \forall i \\ & -y_{ij} \leq x_i - z_{ij} \leq y_{ij} && \forall i,\ j \in J_i \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\ & \mathbf{y} \geq \mathbf{0} \end{aligned}$$

Given the ith constraint of the nominal problem, let J_i be the set of uncertain coefficients

$\tilde{a}_{ij}$, $j \in J_i$, where $\tilde{a}_{ij}$ belongs to the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$ and a_{ij} is the nominal value. This model is less conservative than Soyster's model (Bertsimas and Sim, 2004).

The idea of using "budgets of uncertainty" is introduced by Bertsimas and Sim (2004), as naturally, not all uncertain parameters take their worst-case values. They propose a row-wise uncertainty, meaning that all uncertain parameters in a constraint are independent of those in other constraints. Also, they suggest that the total number of uncertain parameters deviated from their nominal values, should not exceed a predetermined budget of uncertainty. The budget of uncertainty provides the flexibility of deciding on a trade-off between robustness and performance and reduces the conservatism. They also show that the robust counterparts of the LP remain LP, with additional variables and constraints. Bertsimas and Sim (2003) also extend their methods for discrete optimization and network flow problems in a tractable way. They demonstrate that the robust counterpart of an integer program (IP) remains an IP. The nominal formulation is as follows:

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned}$$

Bertsimas and Sim (2004) propose a parameter Γ_i to control the level of robustness against the level of conservatism. Γ_i, takes value in the interval $[0, |J_i|]$. They find the following formulation:

$$
\begin{aligned}
\max \quad & \mathbf{c}'\mathbf{x} \\
\text{s.t.} \quad & \sum_j a_{ij}x_j + \\
& \max_{\{S_i \cup t_i | S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \{\sum_{j \in S_i} \hat{a}_{ij} y_i + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} y_t\} \leq b_i \quad \forall i \\
& -y_j \leq x_j \leq y_j \quad \forall j \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{y} \geq \mathbf{0}
\end{aligned}
$$

Then Bertsimas and Sim (2004), by utilizing strong duality, make an equivalent linear formulation of their model. Later, Bertsimas and Thiele (2006) consider the problem of controlling a supply chain in the presence of uncertainty in demand. They present a robust optimization method by employing the concept of the budget of uncertainty in Bertsimas and Sim (2004) to adjust the level of conservatism. They state that their method is numerically tractable, particularly opposed to dynamic multidimensional programming problems in complex supply chains. Mamani et al. (2016) apply robust optimization to a classic single echelon, N-period inventory problem for a single product. The authors extend the work of Bertsimas and Thiele (2006) by considering the uncertainty set both as symmetric and asymmetric sets, and assess the problem by using uncertainty sets that are motivated by the central limit theorem.

Robust optimization has been applied successfully in many fields of applications including inventory and logistics (Hazır et al. (2010)), facility location (Baron et al. (2011)), healthcare (Hanne et al. (2009), Bortfeld et al. (2008)), finance (Fabozzi et al. (2010)), energy (Bertsimas et al. (2013)). See Ben-Tal and Nemirovski (2008), Bertsimas et al.

(2011), and Gabrel et al. (2014) for comprehensive surveys on RO applications.

RO is especially useful in healthcare settings, in which not considering sources of uncertainty could impact patients' health outcome. Some examples of using RO in healthcare problems are in cancer treatment and addressing challenges in hospitals. For example, Bortfeld et al. (2008) present robust optimization in the application to radiation therapy. They consider the uncertainty due to breathing motion and obtain treatment plans for cancer patients. Also, in the context of patients transportation in hospitals, Hanne et al. (2009) employ a robust approach to the problem of patient flow management.

Decision making under uncertainty can be divided into two categories: static and dynamic. In static robust optimization, decision-makers must take decisions before the realization of the uncertain parameter, and no recourse action will be possible after the realization of uncertainty. These decisions are called "here-and-now decisions". In dynamic decision making, decisions are made in several stages. For example, in two-stage decision making, there are two sets of decisions; first, "here-and-now" or "first stage" decisions; second, "wait-and-see" or "recourse" decisions which are decisions that are made after the realization of the uncertain parameter (Neyshabouri and Berg, 2017). An example of dynamic decision making is the work of Neyshabouri and Berg (2017), who address the problem of scheduling patients by incorporating the uncertainties in the duration of surgery and the length-of-stay in intensive care units. They propose a two-stage robust optimization model which allows managers to control the level of risk that they are comfortable with. A column-and-constraint generation method is developed to solve the model and simulations are used to test the quality of the solution.

Our robust optimization approach and those we survey below fall into the first category, static robust optimization. Chen et al. (2011) investigate the problem of managing hospitals beds for emergency and elective inpatients. They propose a distributionally robust optimization approach to determine the quotas. They develop two models based on the fixed and optimized budget of variation subject to "an expected bed shortfall constraint". They tested the models by simulations based on real hospital admission data. There are approaches other than budgets of uncertainty to reduce the level of conservatism. For example, Meng et al. (2015) propose a distributionally robust optimization model for managing elective admissions in a hospital. They present an optimized budget of variation approach to deal with the ambiguous probability distribution and incomplete information of uncertainties. They obtain the maximum level of uncertainty that the system can have without violating the constraint of expected bed shortfall. A second-order conic problem is presented to solve the robust optimization, and the quality of solutions are tested with simulations based on hospital admission data.

In our study, we base our robust optimization approach on the idea of "budgets of uncertainty". This notion provides decision makers the flexibility of controlling the level of conservatism and decide on a trade-off between the system's performance and the level of protection against uncertainty. We propose the robust counterparts of deterministic models by considering that the uncertain demand belongs to a known uncertainty set that can be controlled by a budget of uncertainty. Our formulation can be used in other applications such as other health problems and manufacturing that have priority levels in patients or customers and uncertainty in parameters. To the best of our knowledge, this work is the first work to address multi-period scheduling of multi-priority patients which explicitly considers different wait time targets for each priority and applies robust optimization to incorporate the uncertainty in patient arrivals. However, patient's preference in appointment dates is not considered in our study.

Chapter 3

Problem Definition

In healthcare settings, treating patients within medically acceptable wait times is vital, because excessive wait times may potentially deteriorate health outcomes. Magnetic resonance imaging (MRI) is one of the critical steps in early detection of diseases. MRI patients, based on their priority level, have an acceptable wait time target which if exceeded, may impact the result of treatment and cause health deterioration. MRI patients in Canada are categorized into 4 priority levels based on the urgency of service. Priority 1, the highest priority, needs emergency service in 24 hours, priority 2 is also urgent and should receive MRI service in less than 2 days, while priority 3 is semi-urgent and need to be served within 10 days and priority 4, which is not urgent, can wait up to several weeks, 28 days (CAR, 2013). Reducing wait times for multi-priority health services is a significant challenge for Canada's government and hospitals. As in 2013, Canadian Association of Radiologists has recommended the 90th percentile service rate for medical imaging as the "preferred retrospective measure" within each priority level (CAR, 2013). However, with current practice, these goals are not met, and over 67% of patients exceed their wait time target, and many exceed it by more than 100 days (Jiang, 2016).

The mismatch between the available capacity of MRI and the increasing number of patients who need MRI exams has increased patients' wait times. This problem can be mitigated by two methods: Increasing the capacity and developing a more practical scheduling method. Capacity expansion is achievable through increasing the number of MRI equipment and facilities and extending the operating hours in each day. MRI machines are expensive, and they need area expansion in hospitals as well (Green et al., 2006). Also, there are restrictions on extending the performing hours in each day. Another method to reduce wait time is through an improved scheduling approach. Poor scheduling is "a major source of operational inefficiency and patient dissatisfaction", according to Chakraborty et al. (2010). On the other hand, a practical scheduling method is a primary means of improving resource utilization and reducing patient wait times (Lowery and Martin, 1989). Currently, hospitals apply a strict scheduling policy, which serves patients with higher priority first. In this policy, lower priority patients are not served as long as a higher priority's queue is not exhausted. This policy neglects how long lower priorities have been waiting to receive the service which leads to longer wait time for lower priorities. However, hospitals aim to serve lower priority patients as well, as long as higher priority ones are served within their wait time target and thus are not in danger.

In this thesis, we address the problem of reducing patients' prolonged wait times in a system of multi-priority levels with different wait time targets. We study an advance scheduling problem which is motivated by the scheduling of magnetic resonance imaging (MRI) exams at Ontario's hospitals. First, we develop a model for the problem of scheduling patients with different priorities and wait time targets over a multi-period finite horizon. In our approach, we consider the "indirect wait" or "intradaywait" of patients, which is the interval between the patient's request day for service and the actual day of service (Truong, 2015). There are different penalty costs for patients waiting from different priorities, which are greater for the higher priority, and the goal is to minimize the total wait times of patients while serving at least 90% of patients within their wait time targets.

In the proposed model, we assume that the demand is known before the horizon. Next, the available capacity of MRI exams is allocated among different priorities based on their demands while ensuring fulfilling the desired service level of 90%. Our model addresses the following key question: How many patients in each priority are going to be served in each period. This decision is made at time 0 of the horizon and patients are served based on a first-come, first-serve policy within the same day, given the allocated capacity.

In Section 3.1, we explain the difference between our approach and current practice. Deterministic formulation of the multi-priority patients scheduling problem with different wait time targets is presented in Section 3.2.

3.1. Strict Policy vs. Our Approach

One of the most commonly-used methods in scheduling patients for MRI exams in hospitals is the strict policy. This policy enforces serving higher priority first without considering how long a patient from a lower priority has waited to receive the service. Thus, as long as there are patients of higher priorities in the queue, lower priorities are not served. This strategy leads to prolonged wait times for lower priority patients, stated in Table 1.1, the 90th percentile of patients with priority 4 wait more than 100 days to receive MRI exams. Therefore, in the proposed model, we aim to develop a more practical method to tackle the issue of these prolonged wait times by providing a model that ensures 90% is met for all patients. We consider the service rate of 90%, because the national maximal wait time access target for medical imaging recommends this service rate as the "preferred retrospective measure" within each priority level (CAR, 2013).

3.2. Definitions

In this section, we address the problem of scheduling patients with multiple priorities P with different wait time targets, w_p, over a finite horizon of T periods indexed by $t \in \{1, 2, ..., T\}$, while fulfilling the target of serving over α_p of patients within their wait time target. In each period t in the horizon, demand, d_{pt}, arises from different priority, p. The decision in period t is how to allocate the available capacities in periods $\{t, t+1, ..., T\}$ among different priorities. We use the two terms "period" and "day" interchangeably as in reality a period normally corresponds to a day.

Notations

Two sets of non-negative integer decision variables are defined as:

x_{ptn}: Total number of patients with priority p who are served in period t after waiting for n periods.

I_{ptn}: Total number of patients with priority p who are not served at the end of period t and are waiting for n periods.

Parameters are as follows:

T: Planning horizon, $\quad t \in \{1, 2, ..., T\}$

P: Priority levels, $\quad p \in \{1, 2, ..., P\}$

n: Number of days that a patient is waiting to receive a service. $\quad n \in \{1, 2, ..., T\}$

d_{pt}: Total arrivals of patients with priority p in period t.

C_t: Total available time of MRI services in period t (capacity).

s_p: Service time of patients with priority p.

w_p: Wait-time target of patients with priority p (days).

α_p: Percentage of patients with priority p who are served within their wait time target (Service level).

$h(n,p)$: Penalty cost for delaying patients type p for n days.

3.3. Deterministic Model

The assumptions in the deterministic model are:

- Patients' arrivals are deterministic and known at the beginning of the horizon.
- Patients can be served after the horizon.
- Patients cannot choose between appointment dates.

The deterministic scheduling problem is then formulated as a mixed integer program [M1]:

$$\text{[M1]:} \quad \min \sum_{p=1}^{P} \sum_{t=1}^{T} \sum_{n=1}^{t} h(p,n) I_{ptn} \tag{3.1a}$$

$$\text{s.t.} \quad I_{ptn} = d_{pt} - x_{ptn} \qquad \forall p \in P, t \in T, \; n = 1 \tag{3.1b}$$

$$I_{ptn} = I_{p,t-1,n-1} - x_{ptn} \qquad \forall p \in P, t \in T, \; \forall n \in \{2, ..., t\} \tag{3.1c}$$

$$I_{ptn} \geq 0 \qquad \forall p \in P, t \in T, n \in \{2, ..., t\} \tag{3.1d}$$

$$\sum_{p=1}^{P} \sum_{n=1}^{t} s_p x_{ptn} \leq C_t \qquad \forall t \in T \tag{3.1e}$$

$$\sum_{n=1}^{w_p} \sum_{t=1}^{T} x_{ptn} \geq \alpha_p \sum_{t=1}^{T} d_{pt} \qquad \forall p \in P \tag{3.1f}$$

$$x_{ptn} \geq 0 \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.1g}$$

$$x_{ptn} \text{ is integer} \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.1h}$$

The objective function (3.1a) minimizes the total wait times of patients to receive the service. There are different penalty costs for patients' waiting from different priorities. The penalty cost, $h(p, n)$, is a function of patients' priority and the number of days they wait to receive MRI exam. Index n is counting the interval between the patient's request day for service and the actual day of service. Different penalty costs for priority levels means one day waiting for a patient with higher priority has a greater penalty than that of one day waiting for a patient with lower priority. Therefore, if there is a patient of higher priority who is not served on the day of arrival, the schedule tries to serve him/her as soon as possible. Arrivals, d_{pt}, occur in each period for each priority. Hospitals provide service for some of the patients based on the available capacity of MRI services in each period, C_t, and the rest are delayed to the next periods.

The number of patients who are waiting to receive the service, I_{ptn}, is calculated by two linear equations; Constraints (3.1b) and (3.1c) describe patients who are waiting for 1 day and more than one day, respectively. Constraint (3.1d) ensures not to serve more patients than the number of arrivals, and constraint (3.1e) restricts the total daily number of performed procedures to the available time of MRI services in each period. Additionally, constraint (3.1f) assures that the desired service level, α_p, of serving patients within their wait time targets is achieved.

Considering constraints (3.1b) and (3.1c) together, leads to the following closed-form expression:

$$I_{ptn} = d_{p,t-n+1} - \sum_{i=0}^{n-1} x_{p,t-i,n-i} \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.2}$$

Then we have the following Model [M2]:

$$\text{[M2]:} \quad \min \sum_{p=1}^{P} \sum_{t=1}^{T} \sum_{n=1}^{t} h(p,n) I_{ptn} \tag{3.3a}$$

$$\text{s.t.} \quad I_{ptn} = d_{p,t-n+1} - \sum_{i=0}^{n-1} x_{p,t-i,n-i} \qquad \forall p \in P, t \in T, \forall n \in \{1, ..., t\} \tag{3.3b}$$

$$d_{p,t-n+1} - \sum_{i=0}^{n-1} x_{p,t-i,n-i} \geq 0 \qquad \forall p \in P, t \in T, \forall n \in \{1, ..., t\} \tag{3.3c}$$

$$(3.1e) - (3.1h)$$

Constraint (3.3b) represents the number of patients of priority p at the end of priority t who are waiting n day(s) to receive the service.

In our deterministic model [M2], we assume that all the demands are known with certainty. Whereas, in reality, daily demand variability of different priority for MRI exams can be observed and it is difficult to predict patients' arrivals from each priority in each day (Jiang, 2016). Therefore, the need arises to develop an optimization approach that takes the uncertainty of arrivals into account and remains numerically tractable as the dimensions of the problem increase. In Section 3.4, we develop a robust optimization approach to guarantee the feasibility and optimality of solutions against uncertainty in demands.

3.4. Robust Optimization

In practice, there is a daily variation in patients' arrivals for MRI exams in each priority, so demands are uncertain, and it is hard to obtain the exact distribution. Thus, in this section, we incorporate the uncertainty in demand without making any assumptions on its distribution. We then develop the robust counterparts of our proposed model. In our robust approach, we only need the information of nominal values and total possible variation of demands from their nominal values for each priority p. Thus, patients' arrivals of priority p in period t, $\tilde{d}_{pt}$, are assumed to be uncertain parameters that take values in a known interval centered at nominal values, $\overline{d}_{pt}$, and half length of variation, $\hat{d}_{pt}$, but their exact values are unknown; $\tilde{d}_{pt} \in \left[\overline{d}_{pt} - \hat{d}_{pt}, \overline{d}_{pt} + \hat{d}_{pt}\right]$. The nominal value refers to the value of demand which is more likely to happen. Let $z_{pt} = (\tilde{d}_{pt} - \overline{d}_{pt})/\hat{d}_{pt}$ be the scaled deviation of demand from its nominal value for each priority p in each period t, which takes values in the interval $[-1, 1]$.

In the deterministic formulation [M2], demand uncertainty affects the objective function (3.3a), and constrains (3.3b), (3.3c) and (3.1f). Since we have uncertainty in the objective function, we can rewrite the model by considering an auxiliary variable y_{ptn} and adding the objective function to the constraints. Assuming that the vector of demand d belongs to an uncertainty set U, robust counterpart of our problem is presented as [M3]:

$$
\begin{aligned}
\text{[M3]:} \quad & \min \sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{n=1}^{t} h(p,n) y_{ptn} && && (3.4\text{a})\\
\text{s.t.} \quad & y_{ptn} \geq \tilde{d}_{p,t-n+1} - \sum_{i=0}^{n-1} x_{p,t-i,n-i} && \forall p \in P, \forall t \in T, && \\
& && \forall n \in \{1,...,t\}, \tilde{d}_{pt} \in U_p && (3.4\text{b})\\
& \sum_{i=0}^{n-1} x_{p,t-i,n-i} \leq \tilde{d}_{p,t-n+1} && \forall p \in P, t \in T, && \\
& && \forall n \in \{1,...,t\}, \tilde{d}_{pt} \in U_p && (3.4\text{c})\\
& \sum_{p=1}^{P}\sum_{n=1}^{t} s_p x_{ptn} \leq C_t && \forall t \in T && (3.4\text{d})\\
& \sum_{n=1}^{w_p}\sum_{t=1}^{T} x_{ptn} \geq \alpha_p \sum_{t=1}^{T} \tilde{d}_{pt} && \forall p \in P, \tilde{d}_{pt} \in U_p && (3.4\text{e})\\
& x_{ptn} \geq 0 && \forall p \in P, t \in T, n \in \{1,...,t\} && (3.4\text{f})\\
& x_{ptn} \text{ is integer} && \forall p \in P, t \in T, n \in \{1,...,t\} && (3.4\text{g})
\end{aligned}
$$

where different priority levels p have distinct associated uncertainty sets, U_p. Therefore, a manager can allocate a unique uncertainty set to each priority level, based on his/her preference. Various types of uncertainty sets are introduced in the robust optimization literature. One concern in RO is to consider uncertainty sets, U, which lead to tractable robust counterparts (Bertsimas et al., 2004). If U is ellipsoidal, for example, then robust counterparts can be written as second-order conic programs, while if U is assumed to be polyhedral, then the equivalent formulation can be written as a linear program (Ben-Tal et al., 2009).

We study two types of uncertainty sets: Box uncertainty set and budget of uncertainty which is a type of polyhedral uncertainty set. Robust counterparts of the problem in case of the box uncertainty set is presented in Section 3.4.1. In Section 3.4.2, following the notion defined by Bertsimas and Sim (2004), robust counterparts are proposed by employing a budget of uncertainty, which is less conservative than the box uncertainty. The budget of uncertainty approach allows the adjustment of the level of robustness of the solution. This adjustment provides a trade-off between the model performance and the level of protection against uncertainty (Bertsimas and Sim, 2004).

3.4.1 Box Uncertainty

In box uncertainty, $\tilde{d}_{pt}$ take values in $\left[\overline{d}_{pt} - \hat{d}_{pt}, \overline{d}_{pt} + \hat{d}_{pt}\right]$, where $\overline{d}_{pt}$ and $\hat{d}_{pt}$ are given parameters as nominal value and variation of demands for each priority p. Then, $z_{pt} = (\tilde{d}_{pt} - \overline{d}_{pt})/\hat{d}_{pt}$ is the scaled deviation of nominal values. The uncertainty set for uncertain parameter $\tilde{d}_{pt}$, is as follows:

$$U_p = \{d \in \mathbb{R}^n : \tilde{d}_{pt} = \overline{d}_{pt} \pm z_{pt}.\hat{d}_{pt}, |z_{pt}| \leq 1 \ , \forall p \in P, t \in T\} \tag{3.5}$$

The robust approach is to maximize the right-hand side of constraints (3.4b), (3.4c), and (3.4e) over the set of scaled deviations. Therefore, all the uncertain demand parameters $\tilde{d}_{pt}$ take their largest value in the interval, $\overline{d}_{pt} + \hat{d}_{pt}$. After replacing the upper bound of $\tilde{d}_{pt}$ in [M3], we have [M4]:

$$
\begin{aligned}
\text{[M4]:} \quad & \min \sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{n=1}^{t} h(p,n) y_{ptn} && && (3.6a)\\
\text{s.t.} \quad & y_{ptn} + \sum_{i=0}^{n-1} x_{p,t-i,n-i} \geq \overline{d}_{p,t-n+1} + \hat{d}_{p,i-n+1} && \forall p \in P, t \in T && \\
& && \forall n \in \{1, ..., t\} && (3.6b)\\
& \sum_{i=0}^{n-1} x_{p,t-i,n-i} \leq \overline{d}_{p,t-n+1} + \hat{d}_{p,i-n+1} && \forall p \in P, t \in T, && \\
& && \forall n \in \{1, ..., t\} && (3.6c)\\
& \sum_{p=1}^{P}\sum_{n=1}^{t} s_p x_{ptn} \leq C_t && \forall t \in T && (3.6d)\\
& \sum_{n=1}^{w_p}\sum_{t=1}^{T} x_{ptn} \geq \alpha_p \sum_{t=1}^{T} (\overline{d}_{p,t} + \hat{d}_{p,t}) && \forall p \in P && (3.6e)\\
& x_{ptn} \geq 0 && \forall p \in P, t \in T, n \in \{1, ..., t\} && (3.6f)\\
& x_{ptn} \text{ is integer} && \forall p \in P, t \in T, n \in \{1, ..., t\} && (3.6g)
\end{aligned}
$$

The box uncertainty set leads to a tractable robust counterpart, but it is unlikely that all demands in each period occur at their worst-case values. Thus, this approach provides a relatively conservative solution.

3.4.2 Budget of Uncertainty

In robust optimization with box uncertainty, we assume all demands of priority p in period t can happen at their worst-case values at the same time. However, as much as it is unlikely that all the demands happen at their nominal values, it is also unlikely that they all occur at their worst-case value. Although considering all demands happening at their worst-case values is the safest approach, it leads to an extremely conservative solution which is not necessarily justifiable in practice (Bertsimas and Sim, 2004). Therefore, we aim to adjust the level of conservatism of the solution by employing the notion of the budget of uncertainty, which has the property of achieving a trade-off between performance and robustness.

Following the concept of the budget of uncertainty, only some of the uncertain parameters can deviate from their nominal values to worst-case values. We then define $\Gamma_p = (\Gamma_1, ..., \Gamma_P)$ as the budget of uncertainty vector for the demand of each priority. $\lfloor \Gamma_p \rfloor$ determines the total number of demand parameters of each priority that are allowed to be equal to their highest possible amount over the horizon. We enforce $\sum_{t=1}^{T} z_{p,t} \leq \Gamma_p, \forall p$, which restricts the total number of scaled deviations from the nominal value to be less than the budget of uncertainty, Γ_p. Our goals are to determine the worst-case cost for a given level of uncertainty and to find a robust solution, which is feasible even when $\tilde{d}_{pt}$ changes in the interval. The uncertainty set for demand, $\tilde{d}_{pt}$, is as follows:

$$U_p = \{d \in \mathbb{R}^n : \tilde{d}_{pt} = \overline{d}_{pt} + z_{pt}.\hat{d}_{pt}, |z_{pt}| \leq 1\,, \forall p \in P, t \in T, \sum_{t=1}^{T} z_{pt} \leq \Gamma_p, \forall p \in P\} \qquad (3.7)$$

Nominal values and variations of demand for each priority can be gathered from the experts who have enough information regarding each priority's arrivals. Also, different priority levels can have unique levels of uncertainty, Γ_p. Thus, based on the manager's preference and information, various budgets of uncertainty can be allocated to different priority levels (Neyshabouri and Berg, 2017).

In addition, the value of budget of uncertainty for each priority p is chosen based on the risk-attitude of the decision maker (Neyshabouri and Berg, 2017). A higher budget of uncertainty allows more demand points, $\tilde{d}_{pt}$, to deviate from their nominal values, and leads to a more conservative schedule. Γ_p, which is not necessarily an integer takes values in the interval $[0, T]$. We have the flexibility of controlling the level of conservatism by varying Γ_p in $[0, T]$. $\Gamma_p = 0$ means that all $\tilde{d}_{pt}$ take their nominal value and constraints are equivalent to those in the nominal problem. Likewise, when $\Gamma_p = T$, it means all $\tilde{d}_{pt}$ occur at their worst-case values, and the problem is equivalent to that under box uncertainty.

The robust approach here is to satisfy the worst-case of constraints (3.4b), (3.4c), and (3.4e) over the set of possible scaled deviations. By the definition of uncertainty set U_p, we can substitute the value of $\tilde{d}_{pt}$ and reformulate problem [M3] as [M5]:

$$\text{[M5]:} \quad \min \sum_{p=1}^{P} \sum_{t=1}^{T} \sum_{n=1}^{t} h(p,n) y_{ptn} \tag{3.8a}$$

$$\text{s.t.} \quad y_{ptn} + \sum_{i=0}^{n-1} x_{p,t-i,n-i} \geq (\overline{d}_{p,t-n+1} + z_{p,t-n+1} \hat{d}_{p,t-n+1}) \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.8b}$$

$$\sum_{i=0}^{n-1} x_{p,t-i,n-i} \leq \overline{d}_{p,t-n+1} + z_{p,t-n+1} \hat{d}_{p,t-n+1} \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.8c}$$

$$\sum_{p=1}^{P} \sum_{n=1}^{t} s_p x_{ptn} \leq C_t \qquad \forall t \in T \tag{3.8d}$$

$$\sum_{n=1}^{w_p} \sum_{t=1}^{T} x_{ptn} \geq \alpha_p (\sum_{t=1}^{T} \overline{d}_{p,t} + \max_{\{|\zeta_{pt}| \leq 1, \forall p,t \sum_{t=1}^{T} \zeta_{pt} \leq \Gamma_p, \forall p\}} \sum_{t=1}^{T} \zeta_{p,t} \hat{d}_{p,t}) \qquad \forall p \in P \tag{3.8e}$$

$$x_{ptn} \geq 0 \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.8f}$$

$$x_{ptn} \text{ is integer} \qquad \forall p \in P, t \in T, n \in \{1, ..., t\} \tag{3.8g}$$

where ζ_{pt} is the scaled deviation of uncertain demand from its nominal value. In our model, since we have a continuous uncertainty set for $\tilde{d}_{pt}$, we have an infinite number of constraints. To eliminate maximizing $\sum_{t=1}^{T} \zeta_{p,t}\hat{d}_{p,t}$ in constraint (3.8e), we need to consider the inner-maximization problem for each priority p separately as follows:

$$\forall p \in P, \quad \text{[M6]:} \quad \max \sum_{t=1}^{T} \zeta_{p,t}\hat{d}_{p,t} \tag{3.10a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} \zeta_{p,t} \leq \Gamma_p \tag{3.10b}$$

$$0 \leq \zeta_{p,t} \leq 1 \qquad \forall t \in T \tag{3.10c}$$

In [M6], the budget of uncertainty, Γ_p is defined for each priority level p. These auxiliary problems arise from maximizing $z_{p,t}\hat{d}_{p,t}$ in constraints (3.4b) and (3.4c), and $\sum_{t=1}^{T} \zeta_{p,t}\hat{d}_{p,t}$ in constraint (3.4e). Since the optimal ζ_{pt} in [-1,1] obtained from constraints that have uncertain demand are opposite to each other, there is no feasible demand in the interval that can satisfy all. Therefore, as a result, we have constraint (3.10c).

Bertsimas and Sim (2004) in their paper, "the price of robustness", define the uncertainty to be row-wise, which means all uncertain parameters that belong to a constraint are independent of those in other constraints. Then by applying the duality theorem, they reformulate the robust counterpart as a linear program. However, in our model, the uncertain demand parameters in constraint (3.4e) are dependent on the uncertain demand in constraints (3.4b) and (3.4c). The linear programming problems [M6] for all p are feasible and bounded, and we can use the strong duality theorem to substitute the inner-maximization problem in constraint (3.8e) with its dual and reformulate the robust counterparts as a single minimization problem, but we need z_{pt} in other constraints (3.8b) and (3.8c).

In order to solve this min-max problem, we first solve auxiliary linear programming problems [M6] for each priority p and find the optimal objective values u_p^* for each problem. u_p^* is the maximum possible amount of uncertainty for each priority p subject to a given Γ_p over the horizon. Then, by adding a new set of constraints into Model [M3], we make sure that the total amount of uncertainty for each priority p over the horizon is equal to u_p^*. Additionally, constraints (3.10b) and (3.10c) are added. By substituting the value of $\tilde{d}_{pt}$ with $\overline{d}_{pt} + z_{p,t}.\hat{d}_{p,t}$ and considering the scaled deviation z_{pt} as a new decision variable, we have [M7] as follows:

$$
\begin{aligned}
\text{[M7]:} \quad & \min \sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{n=1}^{t} h(p,n) y_{ptn} && && (3.11a)\\
\text{s.t.} \quad & y_{ptn} + \sum_{i=0}^{n-1} x_{p,t-i,n-i} \geq (\overline{d}_{p,t-n+1} + z_{p,t-n+1}\hat{d}_{p,t-n+1}) && \forall p \in P, t \in T, && \\
& && \forall n \in \{1,...,t\} && (3.11b)\\
& \sum_{i=0}^{n-1} x_{p,t-i,n-i} \leq \overline{d}_{p,t-n+1} + z_{p,t-n+1}\hat{d}_{p,t-n+1} && \forall p \in P, t \in T, && \\
& && \forall n \in \{1,...,t\} && (3.11c)\\
& \sum_{p=1}^{P}\sum_{n=1}^{t} s_p x_{ptn} \leq C_t && \forall t \in T && (3.11d)\\
& \sum_{n=1}^{w_p}\sum_{t=1}^{T} x_{ptn} \geq \alpha_p(\sum_{t=1}^{T} \overline{d}_{p,t} + u_p^*) && \forall p \in P && (3.11e)\\
& \sum_{t=1}^{T} z_{p,t} \leq \Gamma_p && \forall p \in P && (3.11f)\\
& 0 \leq z_{p,t} \leq 1 && \forall p,t && (3.11g)\\
& x_{ptn} \geq 0 && \forall p \in P, t \in T, &&
\end{aligned}
$$

$$\forall n \in \{1, ..., t\} \qquad (3.11h)$$

$$x_{ptn} \text{ is integer} \qquad \forall p \in P, t \in T, \; \forall n \in \{1, ..., t\} \qquad (3.11i)$$

where u_p^* is the optimal objective function of auxiliary problems [M6], and z_{pt} is a decision variable. Considering z_{pt} as a decision variable provides the model the flexibility of spreading the total amount of uncertainty over the horizon. u_p^* in constraint (3.11e) maintains the worst-case deviation of demands from their nominal values. Other constraints and parameters are the same as those in [M5].

Chapter 4

Results

In this chapter, the proposed robust approach and the non-robust scheme are implemented on multiple demand datasets. Then, the results are analysed to investigate the effectiveness of the robust model in minimizing wait times while maintaining the desired service level under existing uncertainties in demands. Based on the given demand datasets, we solve each model to determine a schedule which is later used to compute wait times and service levels for each priority over different demand instances. In Section 4.1, the parameters and demand instances are explained. Sections 4.2, 4.3, and 4.4 discuss the model parameters. Sections 4.4.1, 4.4.2, and 4.5 provide some more detail on the implementation of the proposed scheduling scheme. In Section 4.6, we perform sensitivity analysis for the different parameters of the robust model to study their influence on the average wait times, service level, and capacity allocation. Our proposed approach and strict policy are compared in terms of priority's wait time in Section 4.7. Finally, in Section 4.8, scheduling from the robust approach and non-robust model are applied to other datasets. Two metrics, service level and patients average wait time for each priority, are calculated to assess the performance of robust compare to non-robust model after the realization of uncertain demands.

4.1. Data

In this research, two groups of datasets are used for the numerical testing. In the first group, parameters are taken from Jiang (2016) which are based on the MRI admission data of 74 hospitals in Ontario. The daily number of patient's arrivals are generated based on the normal distribution with parameters of $\mu = 41.30$ and $\sigma = 12.81$. Three different scenarios are presented based on various percentages of patients with different priorities. We then generate 10 random datasets of the aggregate number of arrivals in each period for each scenario. In a particular hospital, among all the patients in need of an MRI exam, a few are of priorities 1 and 2 (2% and 7%) while the demand for the patients with the priorities of 3 and 4 are higher, i.e., 14% and 77% respectively, where these percentages show the proportions of patients in each priority. This dataset is used in Section 4.8 to compare RO and the non-robust approach in terms of service level and wait times. The datasets are explained in detail in Section 4.8.1.

Since the demand levels for priorities 1 and 2 are low, to show the performance of the proposed approach and the impact of important parameters, another group of data is generated. The new group is generated based on the uniform distribution with a higher minimum and maximum value of demand. The aim of generating this group of datasets is to have higher and relatively equal demands in each priority. One example of these datasets is summarized in Table 4.1. We use these datasets in the analysis of Sections 4.3-4.7.

Parameters are as follows: The planning horizon, T, is a finite horizon consisting of

365 periods (days) which is based on the fiscal year. A fiscal year refers to a period that Canada's Ministry of Health uses for its regulations. In a medical calendar, a fiscal year is from 1 January to 31 December. Therefore, to consider the desired service level of patients in regulations, we need to consider a year as the time horizon. Hospitals serve four priorities, p of patients with different wait time targets, w_p for MRI services. Wait time targets for patients with priority 1, 2, 3, and 4 is 1 day, 2 days, 10 days, and 28 days, respectively. These priorities are the outcomes of physician assessments and related wait time targets are determined by Canadian association of radiologists in collaboration with the Ontario Government (WTA, 2013). Uncertain demand for each priority belongs to an interval which is centered by the nominal demand, $\bar{d}_{pt}$ and the half-length of variation of demand, $\hat{d}_{pt}$. (Nominal value refers to the value of demand which is more likely to happen.)

In our numerical analysis, we assume that the service time for each priority, s_p, is constant and takes the same value for all priorities, but this assumption can be relaxed. The desired service level for each priority is to serve at least 90% within their wait time targets, where it is based on the health regulation of Canada's ministry of health. Other parameters in Table 4.1, like budget of uncertainty, Γ_p, capacity of MRI services in each period, C_t and penalty costs, $h(p, n)$ are explained in Sections 4.2, 4.3 and 4.4, respectively.

$p = 4$	Number of priorities of patients
$T = 365$	Planning horizon
$h_1 = 4, h_2 = 3, h_3 = 2, h_4 = 1$	Penalty costs of delaying priorities 1, 2, 3, and 4
$s_p = 1,\ \forall p$	Service time of each priority
$w_1 = 1, w_2 = 2, w_3 = 10, w_4 = 28$	Wait time targets of priorities 1, 2, 3, and 4
$c_t = 110,\ \forall t$	Capacity of MRI services in each period
$\overline{d}_{1t} = 20, \hat{d}_{1t} = 10,\ \forall t$	Nominal value and variation of demand for priority 1
$\overline{d}_{2t} = 30, \hat{d}_{2t} = 14,\ \forall t$	Nominal value and variation of demand for priority 2
$\overline{d}_{3t} = 25, \hat{d}_{3t} = 12,\ \forall t$	Nominal value and variation of demand for priority 3
$\overline{d}_{4t} = 28, \hat{d}_{4t} = 13,\ \forall t$	Nominal value and variation of demand for priority 4
$\Gamma_p = 100,\ \forall p$	Budget of uncertainty for each priority

Table 4.1: Sample of input data

4.2. Budget of Uncertainty

In the proposed robust approach, we assume that the total number of uncertain demand points in each priority that can deviate from their nominal values to their worst-case realizations over the finite horizon is constrained to be less than a predetermined budget of uncertainty. The budget of uncertainty allows managers to control the level of conservatism in the system. It can take the values of 0 to 365. In the case of $\Gamma = 0$, there is no uncertainty in the system and all demands happen at their nominal values while in the case of $\Gamma = 365$, all demand points occur at their worst-case realizations. The budget of uncertainty is defined for each priority, in case a manager wants to consider higher uncertainty in a specific priority level. The total amount of uncertainty in each priority over the horizon is predetermined, while the proposed robust approach has the flexibility of spreading the given uncertainties over periods.

4.3. Minimum Capacity Required for a Level of Uncertainty to Satisfy 90% Service Level

As mentioned in previous chapters, Canada's Ministry of Health's regulation is set to serve at least 90% of patients in each priority within their wait time targets. However, with current scheduling policies and capacities, hospitals are not able to achieve this goal (see Table 4.1 and Jiang (2016)). The limited number of MRI machines and the daily operating hours are main obstacles to meet this target. Besides, the existing uncertainty in demand increases the mismatch between demands and capacity in each day. Thus, one necessary analysis is to find the minimum capacity that is required to achieve the desired service level of 90% under the presence of uncertainty in demands. Also, because the minimum required capacity depends on the amount of uncertainties in the system, we need to take different values of the budget of uncertainties into account as well. Therefore, we solve the robust model for the different budget of uncertainties while keeping all other input values constant, including the demand of each priority in each day and service level. Then for each budget of uncertainty we determine the minimum required capacity which provides a feasible solution. Figure 4.1 presents the level of uncertainty corresponding to a given capacity to meet the service level of 90%.

Figure 4.1 shows that the minimum required capacity when $\Gamma = 0$ is 93 hours, while it is 137 hours in case of $\Gamma = 365$. Thus, if all patient demand points happen at their nominal values, 93 hours of MRI is enough to serve at least 90% of patients from different priorities within their wait time targets. However, if the worst-case scenario happens, which means all patients demand occur at their maximum values, to achieve the desired level of service, the manager needs to increase the capacity by at least 44 hours to 137 hours. We assumed that capacity and service time are integer-valued and based on hours; however, shorter

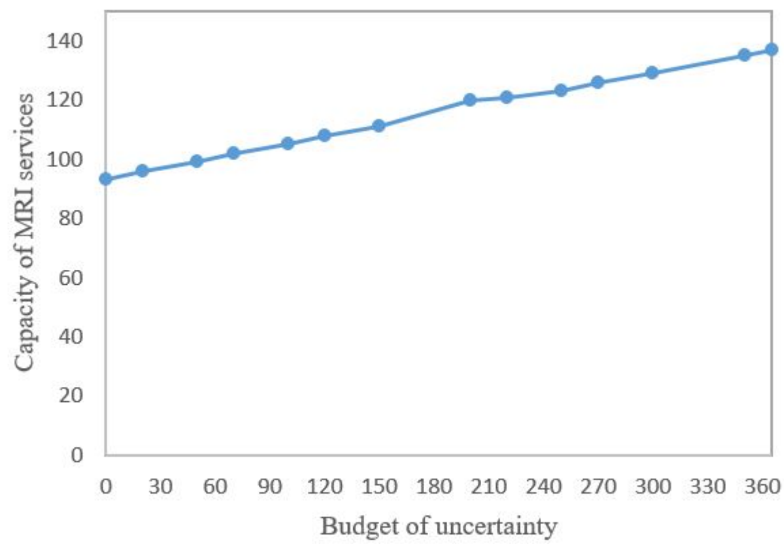


Figure 4.1: Relationship between budget of uncertainty and capacity

time periods can be considered as well.

This analysis also provides a useful tool for managers to decide on how much uncertainty they want to consider in their systems based on the available capacity or decide on whether to increase the number of MRI machines or the daily operating hours to fulfill the desired service rate in the presence of higher uncertainty.

Note, for the rest of the results, we assume that we have at least minimum level of capacity, C_t, in each period for the given budget of uncertainty, Γ_p.

4.4. Penalty Function

In healthcare systems, delaying patients from different priority level may have different outcomes. In case of priorities 1 and 2, delaying service can deteriorate patients' health. Thus, in our proposed scheduling approach, an important factor to manage the queue is to consider different penalty costs for delaying patients from different priorities. Penalty cost means how much cost would incur if a patient waits for one more day in the system. First, we investigate the impact of considering different penalty costs on priority's service level and wait times in Section 4.4.1. Then, allocated capacity to each priority is analysed under two cases of considering the same and the different penalty costs for each priority level in Section 4.4.2.

4.4.1 Effect of penalty cost on service level and wait time

In this section, we analyse how considering different penalty costs for different priorities influence priority's service level and average wait time. The average service level and average wait time of 10 random datasets for each case are summarized in Table 4.2. All results are obtained by keeping all other input parameters constant, including $\Gamma_p = 100$ and $C_t = 107$ units in each period. In Table 4.2, the column of penalty cost consists of four cases of different waiting cost. In the first case, it is assumed that there is no difference between delaying patients among different priority levels, whereas in the rest of the cases, the longer the wait time target, the lower penalty is applied on waiting patients. In case 2, higher priorities have higher penalty of waiting, $h_1 = 4$, $h_2 = 3$, $h_3 = 2$, and $h_4 = 1$. In case 3, the values of penalty costs for each priority are set relative to their different wait time target, $h_1 = \frac{28}{1}$, $h_2 = \frac{28}{2}$, $h_3 = \frac{28}{10}$, and $h_4 = \frac{28}{28}$, where 28 is the highest wait time target which related to priority 4. Finally, in case 4, the waiting cost is the function of waiting days and priorities; that means that the penalty cost of a patient with higher priority is extremely greater than that for a lower priority patient if they wait longer in the system to receive the service. The weights for each priority is the same as case 3.

In Table 4.2, there are two numbers under the third, fourth, fifth and sixth columns; the

Case	Penalty cost	Average number of waiting days, Service level (%) (α_p)				Average wait time for each priority, ($\bar{wt}_p$)			
		P1	P2	P3	P4	wt_1	wt_2	wt_3	wt_4
1	$h_1, h_2, h_3, h_4 = 1$	1, 92.7	1.7, 90	10, 90	16, 92.2	1.6	2.3	3	2.5
2	$h_1 = 4, h_2 = 3, h_3 = 2, h_4 = 1$	1, 96.7	1, 94.5	6.7, 92	28, 90	0.19	0.5	1.4	7.7
3	$h_1 = 28, h_2 = 14, h_3 = 2.8, h_4 = 1$	1, 97.6	1, 94.8	7, 90.9	28, 90	0.12	0.6	1.4	7.8
4	$h_1 = 28n, h_2 = 14n, h_3 = 2.8n, h_4 = 1$	1, 95.2	1, 93.9	4.7, 91.6	27.7, 90	0.23	0.5	1.2	8.2

Table 4.2: Impact of penalty costs on service level and wait times

First number shows the number of waiting days in which the service level exceeds 90% and the second number is priority's service level, α_p. The service level refers to the percentage of patients who received an MRI exam within their wait time target, and the percentage of patients who exceeded wait time target can be calculated as $1 - \alpha_p$. The next four columns show the average wait times of patients in each priority (days). In a practical scheduling, we expect a lower number of waiting days and a higher service level for higher priorities.

Comparing case 1 with the other cases, we observe that by placing a higher penalty cost for a higher priority, the service level of the higher priority is significantly increased and patients with lower priority are delayed within their wait time targets while achieving the 90% service level. As shown in rows 2-4, there are trade-offs between the service levels and wait times of each priority. For example, in case 3, by promoting the higher priority, the percentage of priority 1 patients who exceed their wait time target is reduced by 0.9% compared to that in case 2. This reduction is in exchange for delaying patients' priorities 2, 3 and 4 for slightly higher times. Also, in case 4, patients with priorities 2 and 3 can receive the service sooner, while fewer patients with higher priorities 1 and 2 exceeded their wait time target to receive the service.

4.4.2 Effect of penalty cost on allocated capacity

In this section, the effect of penalty cost on the percentage of allocated capacity among priorities is analysed. In Figure 4.2, the case of having the same penalty costs for different priorities is considered. As we observe, at the first periods, capacity is allocated based on patients' arrivals. However, approaching the end of the horizon, due to the same penalty cost for delaying different priorities, the schedule doesn't differentiate between delaying higher priorities and lower ones. Capacities in each period are allocated only in a way to meet serving at least 90% of patients within their wait time targets.

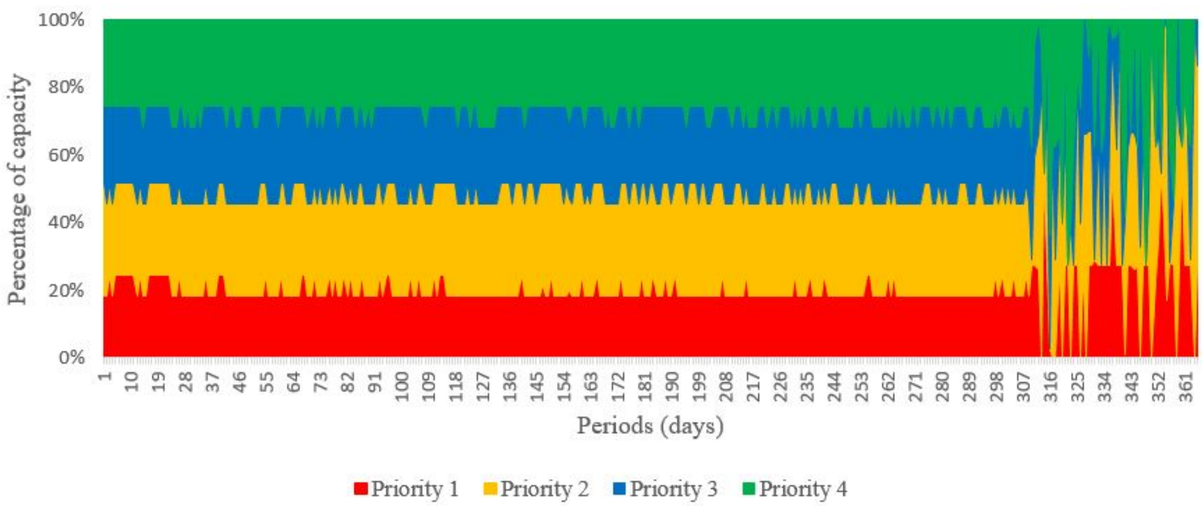


Figure 4.2: Percentage of capacity allocated to each priority per day, $\Gamma_p = 100$, $C_t = 110$

On the other hand, Figure 4.3 shows the allocated capacity among priorities in case of considering different penalty costs for different priorities. In this analysis, the penalty costs are assumed as $h_1 = 4$, $h_2 = 3$, $h_3 = 2$, and $h_4 = 1$. As we observe in Figure 4.3, the allocated capacity per day depends on the patients' arrival in each period. When it approaches the end of the horizon, the schedule allocates more capacity to priority 1, 2 and 3, and delays priority 4 to after the end of the horizon within their wait time targets.

This behavior can occur due to a few reasons. First, the goal is to serve at least 90% of each priority within their wait time targets. Second, in the objective function, we consider a higher penalty cost for delaying the higher priority and a low penalty cost for delaying lower priorities. Therefore, in comparison to Figure 4.2, the obtained schedule differentiates among priorities based on penalty cost as well. Also, the schedule tends to delay patients of priority 3 and 4 after the end of the horizon within their wait time target, since they have longer wait time targets. The observed behavior is also partially due to the difference in spreading the total uncertainty over periods (see Figure 4.4). As observed, uncertainties are mostly spread in the last few periods. The total amount of uncertainty in each day is defines as $\sum_{p=1}^{P} z_{pt}\hat{d}_{pt}$ where $\sum_{t=1}^{T} z_{p,t} \leq \Gamma_p$ and $0 \leq z_{p,t} \leq 1$, $\forall p \in P, t \in T$.

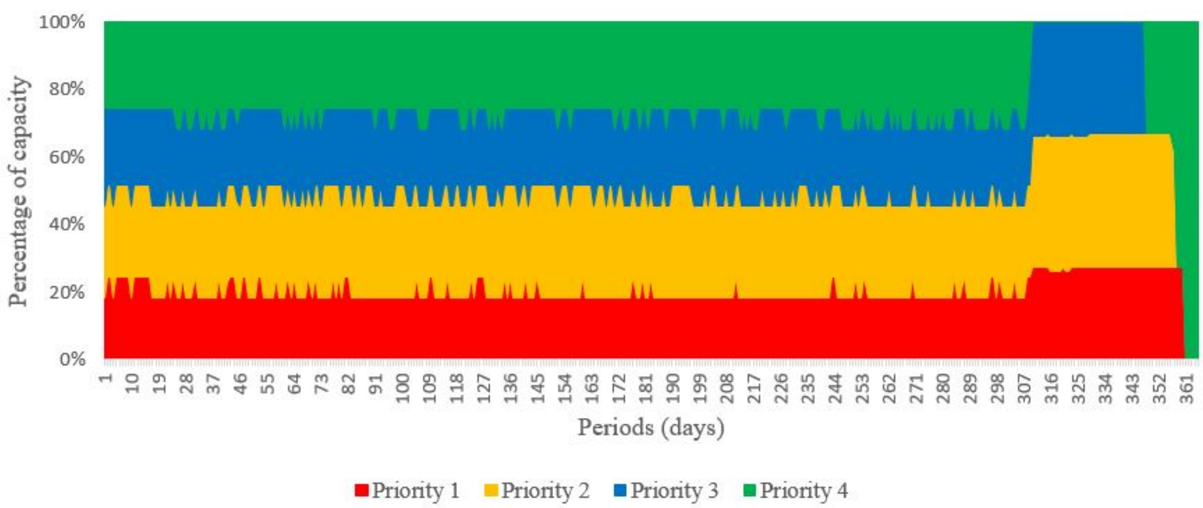


Figure 4.3: Percentage of capacity allocated to each priority per day, $\Gamma_p = 100$, $C_t = 110$

Concluding from Table 4.2 and Figures 4.2 and 4.3, applying higher penalty cost for delaying patients from higher priority enhances the service level of the higher priority patients and places more emphasis on allocating capacities to them, compared to the lower priority in which exceeding wait time target has lower importance. In addition, a manager

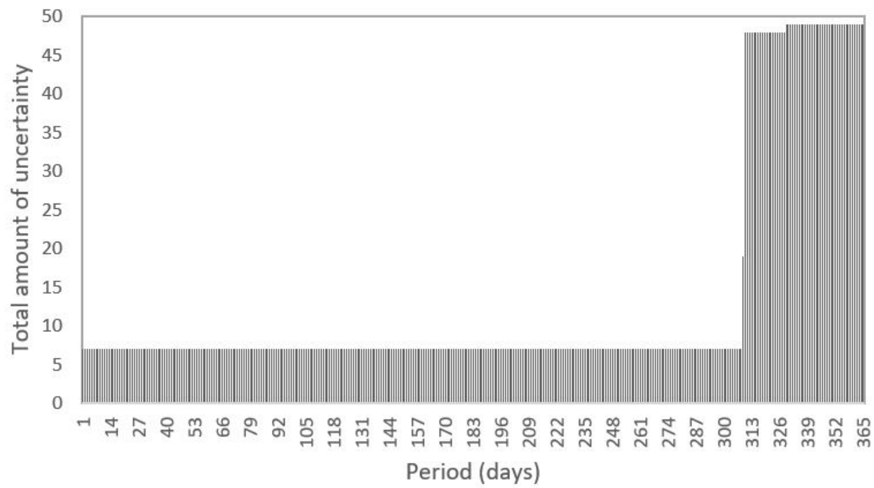


Figure 4.4: The total amount of uncertainty in each day, $\Gamma_p = 100$

can decide on the different values and the function of penalty costs based on his/her preference for each priority. In the rest of the analysis, we assume the penalty costs of case 2 in Table 4.2.

4.5. Effect of Terminal Cost on Allocated Capacity

One of the drawbacks of schedules mentioned in Section 4.4.2 is that there is no penalty for not serving patients at the end of the horizon because of the finite horizon. The schedule tends to delay some patients after the end of horizon within their wait time targets, especially priorities 3 and 4 patients since their wait time targets are longer. Thus, the capacities at ending periods are used mostly to serve priority 1, 2 and 3. To avoid this issue, we analyze the impact of placing a penalty (terminal) costs for patients who are not served by the end of the horizon. As we observe in Figure 4.5, by considering terminal costs, which are higher for not serving higher priorities within the horizon, the schedule forces to allocate more MRI capacity to patients with higher priorities at the first periods of the horizon to prevent higher terminal costs. Also, the higher demand during the beginning periods is due to the spreading of uncertainty. Figure 4.6 displays the total uncertainty which is spread mostly in early periods, and as a result, the demand is taking its maximum value in the uncertainty set.

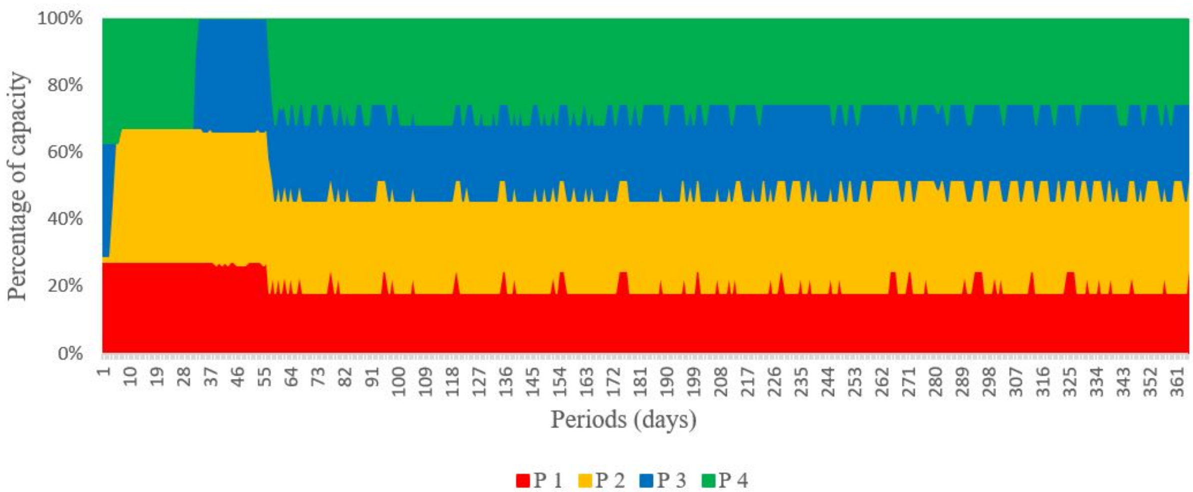


Figure 4.5: Percentage of capacity allocated to each priority per day in the presence of terminal cost, $\Gamma_p = 100$, $C_t = 110$

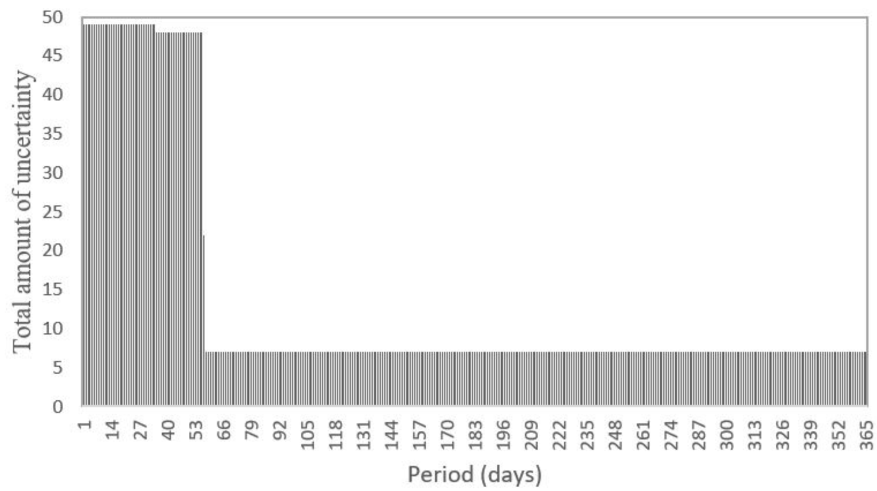


Figure 4.6: The total amount of uncertainty in each day, $\Gamma_p = 100$

4.6. Sensitivity Analysis

In this section, we perform sensitivity analysis over some of the important parameters of the model to see the effect of each one on the scheduling, wait time targets, average wait time and service level. We also introduce some analysis for managerial decisions about the required capacity of MRI services, the number of MRI machines and the amount of uncertainty that can be considered in the system.

It is important to note that in all the results of the current section if the situation is not explicitly specified, we have assumed that we are meeting the regulation of serving at least 90% of patients in each priority within their wait time targets.

4.6.1 Service level versus minimum capacity required

Hospitals' target is to achieve the desired level of service; i.e., to serve over 90% of patients from each priority within their wait time target. However, with the current scheduling policies and the amount of available MRI services, this target is not met. Thus, determining the minimum level of capacity to achieve the desired goal becomes essential. Moreover, it is useful for managers to know what is the minimum required capacity for the desired service level. Therefore, in this section, we study the relationship between service level and capacity. We solve the robust model for a given dataset of demand and budget of uncertainty, $\Gamma_p = 100$. Then, for different amounts of service level, we find the minimum capacity which guarantees a feasible solution. The relationship between service level and capacity is shown in Figure 4.7.

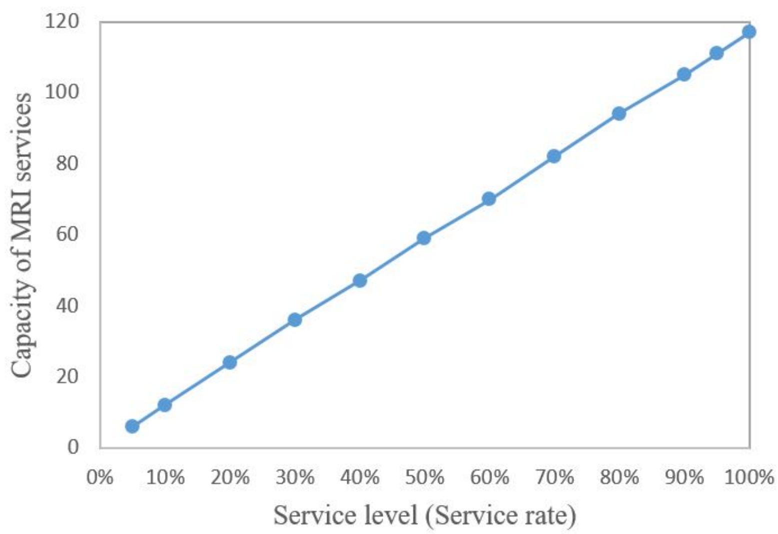


Figure 4.7: Minimum capacity required to achieve desired service level

As shown in Figure 4.7, for this dataset, to achieve the service level of 90%, the hospital needs to have at least 105 hours of MRI services and to achieve, for example, the service level of 50%, it requires 59 hours of MRI services. This analysis provides a useful tool for managers to decide on capacity expansion to achieve the desired service level. The increase in capacity can be obtained by increasing the number of MRI machines or increasing the daily operating hours.

4.6.2 Effect of capacity on wait time

Figure 4.8 shows the number of days that patients in each priority need to wait to receive service. In this analysis, the available capacity of MRI services is 105 hours, and the budget of uncertainty is 100 for each priority type. We observe that the schedule is treating 90% of patients within their wait time target while delaying the other 10% to after their wait time targets. As noted in Figure 4.8, among this 10%, some patients with priority 1 with wait time target of 1 day are waiting as much as 26 days to get the service. The length of wait times reaches 32, 60, and 93 days for patients with priority 2, 3, and 4, respectively. This behavior is caused by the limitation on the daily performed procedures of MRI and the number of daily arrivals.

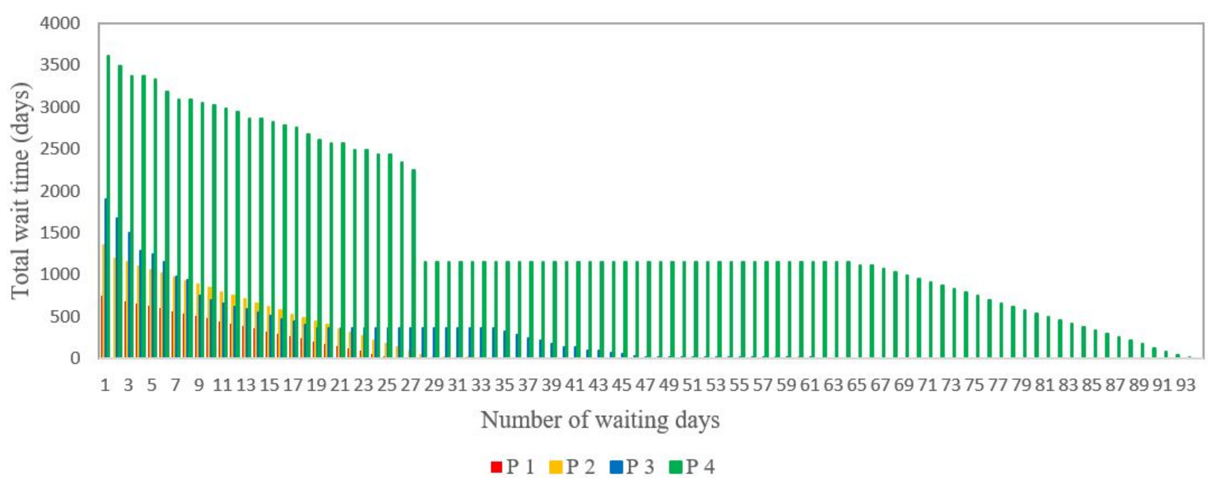


Figure 4.8: Wait time of patients in each priority who are waiting n days to get service, $C_t = 105$, $\Gamma_p = 100$

One way to reduce the length of wait times is through enhancing the capacity. Since

MRI machines are expensive and also there are some constraints on daily procedures, increasing capacity is not an easy way. Therefore, an analysis is needed to see the effect of increasing capacities on the performance. Figure 4.9 displays the impact of increasing the daily capacity by 5 hours and having the capacity of 110 hours per day. As shown in Figure 4.9, the length of wait times for priority 1, 2, 3, and 4 are reduced to 6, 9, 20, and 57, respectively.

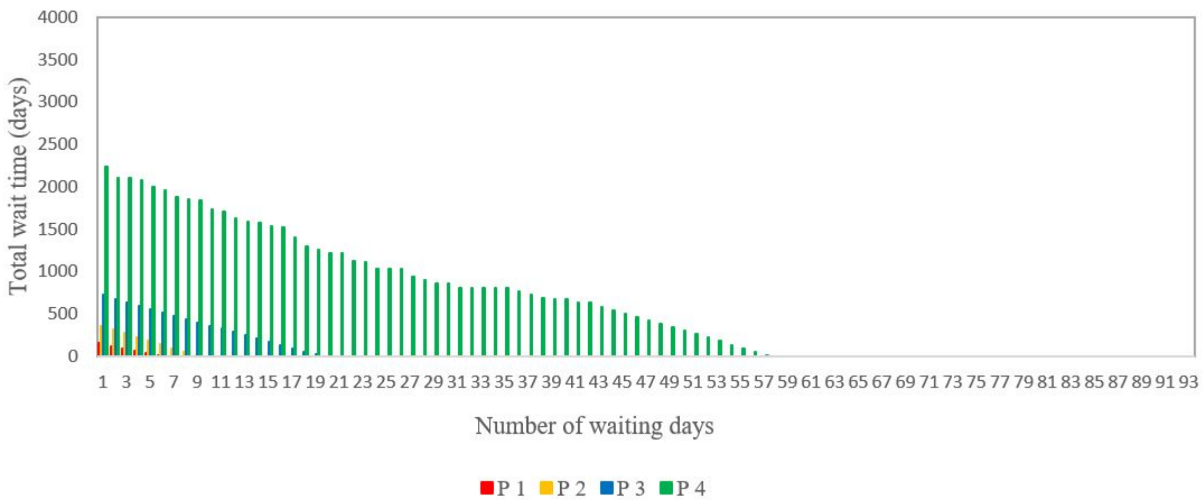
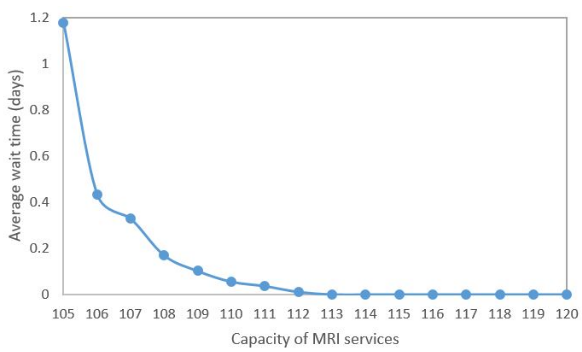


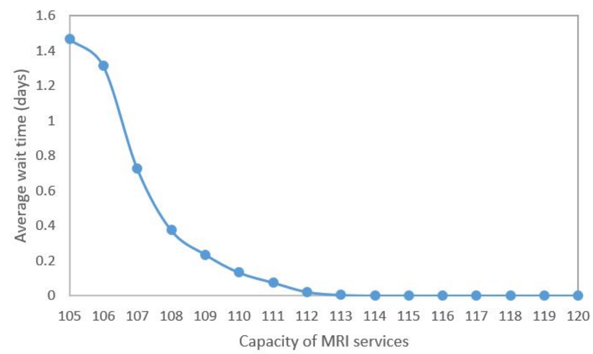
Figure 4.9: Wait time of patients in each priority who are waiting n days to get service, $C_t = 110$, $\Gamma_p = 100$

Also, as noted in Figures 4.8 and 4.9, by increasing the capacity from 105 to 110 hours per day, the total wait time of patients in each priority is reduced. For example, when the capacity is 105 hours per day, the total wait times (days) of patients with priority 4 who have waited for 29 days which exceeded their wait time target of 28 days, is 1152 days while it is 856 days when capacity is increased by 5 hours per day.

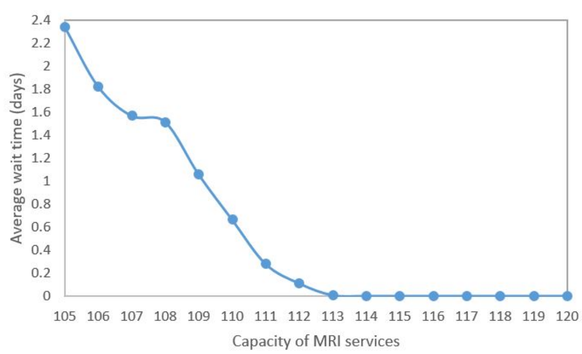
Figure 4.10, shows how the change in capacity influences the average wait time of each priority and average wait time of patients. Results are obtained by changing capacity while keeping all other inputs constant. We observe that wait times are sensitive to changes in



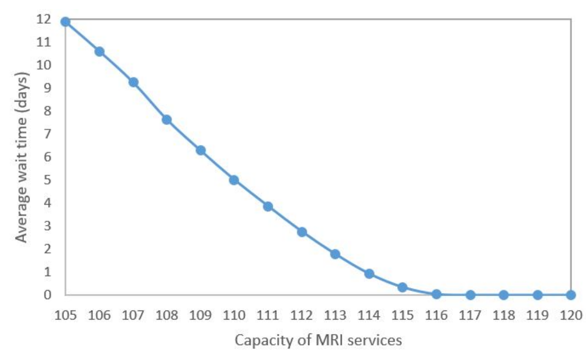
(a) Average wait time of priority 1 (days)



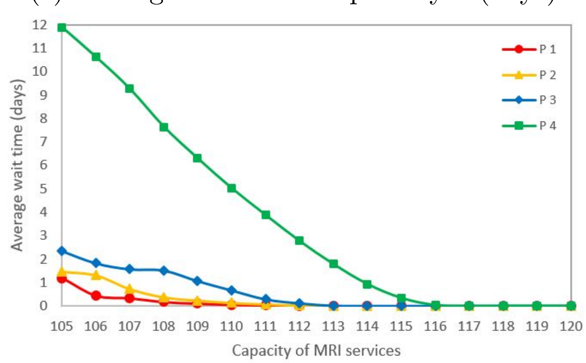
(b) Average wait time of priority 2 (days)



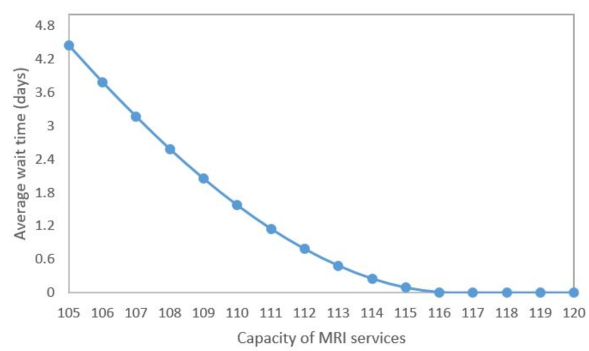
(c) Average wait time of priority 3 (days)



(d) Average wait time of priority 4 (days)



(e) Average wait time of priority p (days)



(f) Average wait time of patients (days)

Figure 4.10: Effect of capacity on average wait time, $\Gamma_p = 100$

capacity. As noted in Figures 4.10(a)-4.10(d), when the capacity increases, the average wait times of all priorities are reduced significantly. In Figure 4.10(a), the minimum required capacity to that priority 1 patients are not delayed is 112 hours, while it increases to 113 hours for priorities 2, and 3, and 116 hours for priority 4. Figure 4.10(e) shows the average wait time of all patients. This analysis provides managers with a tool to decide on expanding the capacity based on the preference for each priority. For example, if all priorities are important to be served without waiting, 116 hours of MRI exams are needed, whereas if only priorities 1 and 2 are vital to serve without waiting, 112 hours is enough. Figure 4.10(f) illustrates the impact of changing capacity on average wait time of patients independent of their priority level. As capacity increases by 1 hour, from 105 to 106 hours, the average wait time of patients is reduced by 1 day.

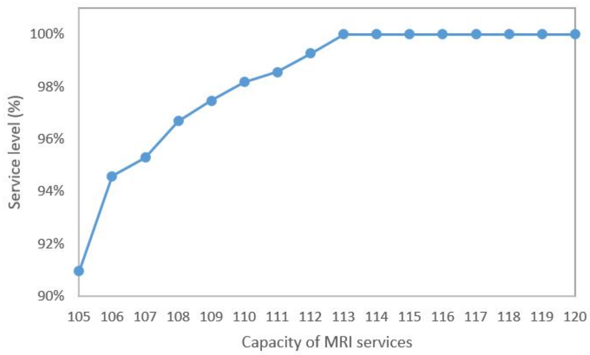
4.6.3 Effect of capacity on service level

In this section, we analyze the impact of changes in capacity on patients' service level. We study the effect of MRI capacity on service levels of each priority as shown in Figures 4.11(a)-4.11(d). In this analysis, the minimum required capacity to meet the service level of 90% is 105 hours for the given $\Gamma_p = 100$, which can also be obtained from Figures 4.1 and 4.7. We observe that when capacity increases from 105 to 106 hours, the extra capacity is used to serve patients of priority 1 and 2, while the service level of patients priority 3 and 4 are not changed. This behavior is because of the higher penalty cost for delaying patients with higher priorities. Thus, in case of capacity expansion, the model prefers to serve priorities 1 and 2, first. When capacity increases to 108 hours per day, more patients priority 3 are served and after 113 hours the service level of patients priority 4 starts to increase significantly.

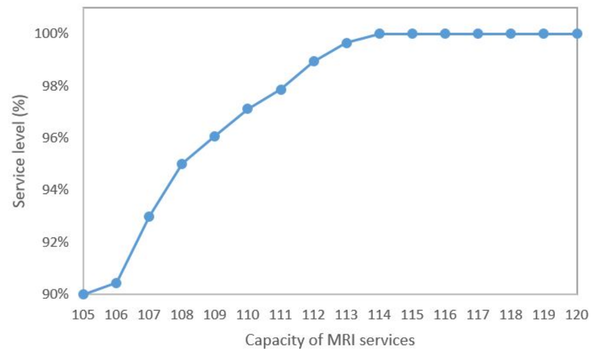
As it is shown in Figure 4.11(e), all patients of priorities 1 and 2 are served when available hours of MRI services is 113 in each day, while patients of priority 3 and 4 are all served when the capacity increase to 114 and 117 hours, respectively. Based on this analysis, decision makers can decide on the expansion of capacity based on their preference for different priority levels. For example, the capacity expansion of 113 hours is sufficient if serving all patients priority 1 is more important while serving over 90% of patients with other priorities. Similarly, Figure 4.11(f) shows that as the capacity of MRI services increases, the weighted average of the service level increases as well. The weighted average of the service level is calculated as $\sum_{p=1}^{P}(\alpha_p \sum_{t=1}^{T} d_{pt}) / \sum_{p=1}^{P} \sum_{t=1}^{T} d_{p,t}$. We observe that by increasing the capacity by 6 hours to 111 hours, the weighted average of the service level increases by 5%. Also, to meet 100% service level, an increase of 12 hours per day is required.

4.6.4 Effect of capacity on number of served patients

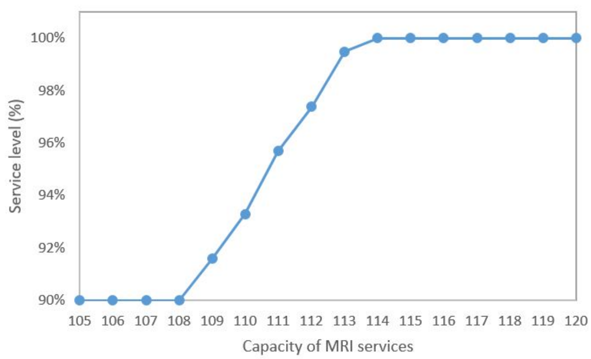
Figure 4.12 shows the number of patients in each priority who are being served after $n-1$ days of waiting, where $n \geq 1$. The available capacity of MRI services in each day is 105 hours for the given $\Gamma_p = 100$. Also, Figure 4.13(a) shows the number of patients in each priority who are served in their day of arrival. It can be seen that patients with priority 1 and 2 are mostly served on their first day of arrival. However, the schedule delays patients with priority 3 and 4 within their wait time targets. 1091 patients priority 4 patients are delayed for 28 days. On the other hand, by increasing the capacity of MRI services by 5 hours per day, not only the number of served patients at the day of arrival is increased (Figure 4.13(b)), but also the number of patients of priority 4 who need to wait for 28 days is reduced to 51 (Figure 4.14).



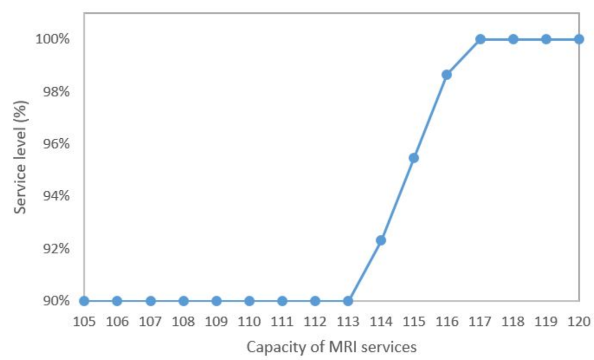
(a) Service level of priority 1



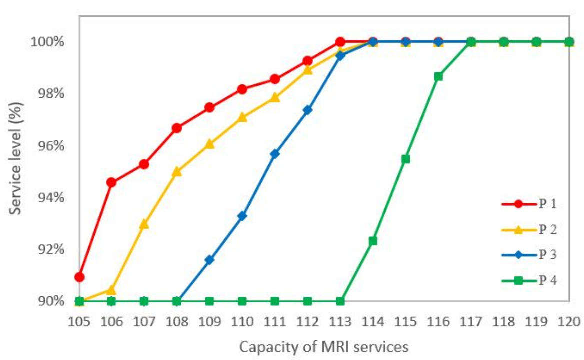
(b) Service level of priority 2



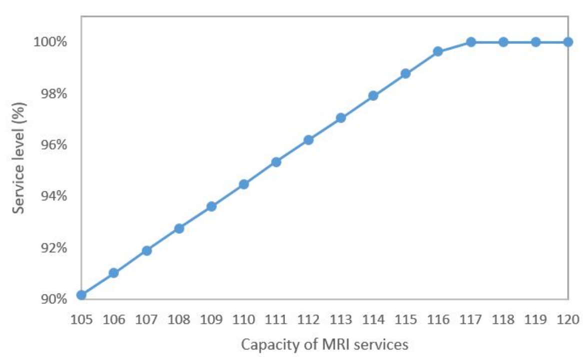
(c) Service level of priority 3



(d) Service level of priority 4



(e) Service level of priority p



(f) Weighted average of patient's service level

Figure 4.11: Effect of capacity on service level, $\Gamma_p = 100$

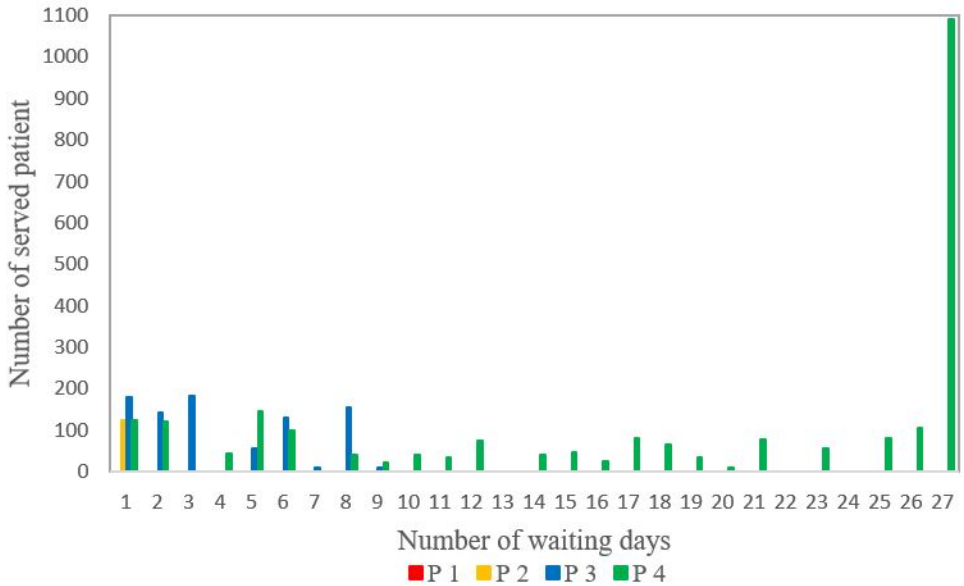
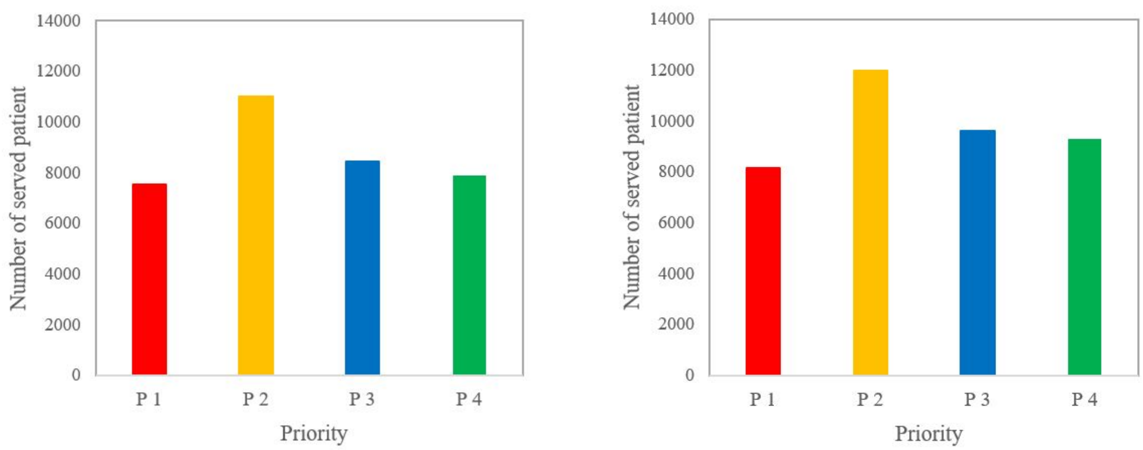


Figure 4.12: Number of patients in each priority who are served after n days of waiting, $C = 105$, $\Gamma = 100$



(a) Number of patients in each priority who are served on the arrival day, $C = 105$, $\Gamma = 100$

(b) Number of patients in each priority who are served on the arrival day, $C = 110$, $\Gamma = 100$

Figure 4.13: Effect of capacity on number of served patients, $\Gamma_p = 100$

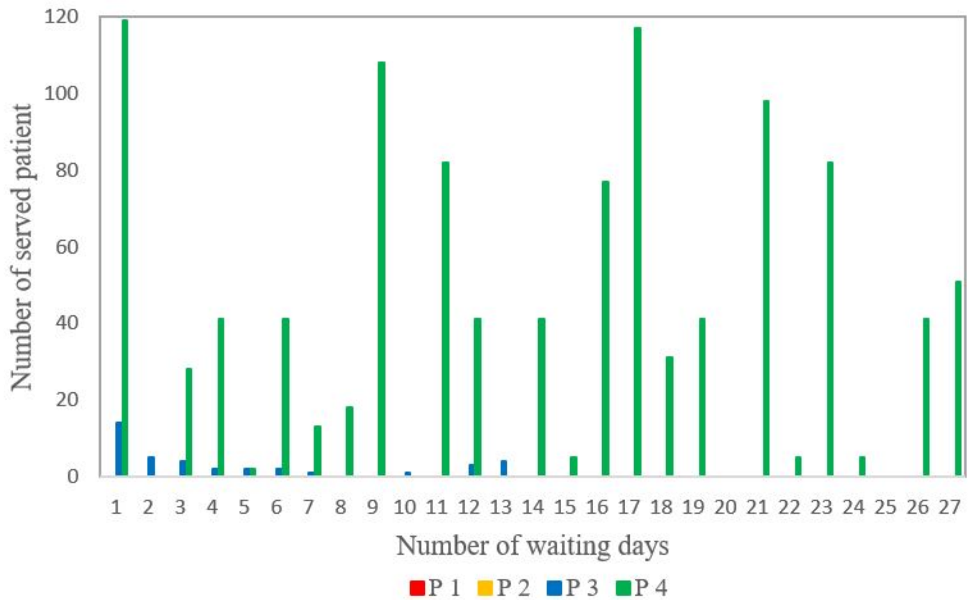


Figure 4.14: Number of patients in each priority who are served after n days of waiting, $C = 110$, $\Gamma = 100$

4.7. Strict Policy

As mentioned in the previous chapters, the strict policy is the common method that is used in hospitals to schedule MRI patients. This policy places emphasis on serving higher priority patients while neglecting the wait time of lower priority patients in the system. Thus the lower priorities are not served until the pool of higher priorities is exhausted. This method leads to prolonged wait time for lower priorities. As it states in Table 1.1, the 90th percentile of patients with priority 4 waits more than 100 days to receive MRI exam. However, often serving some lower priorities before higher priorities is more practical as long as we are serving lower priorities within their wait time target.

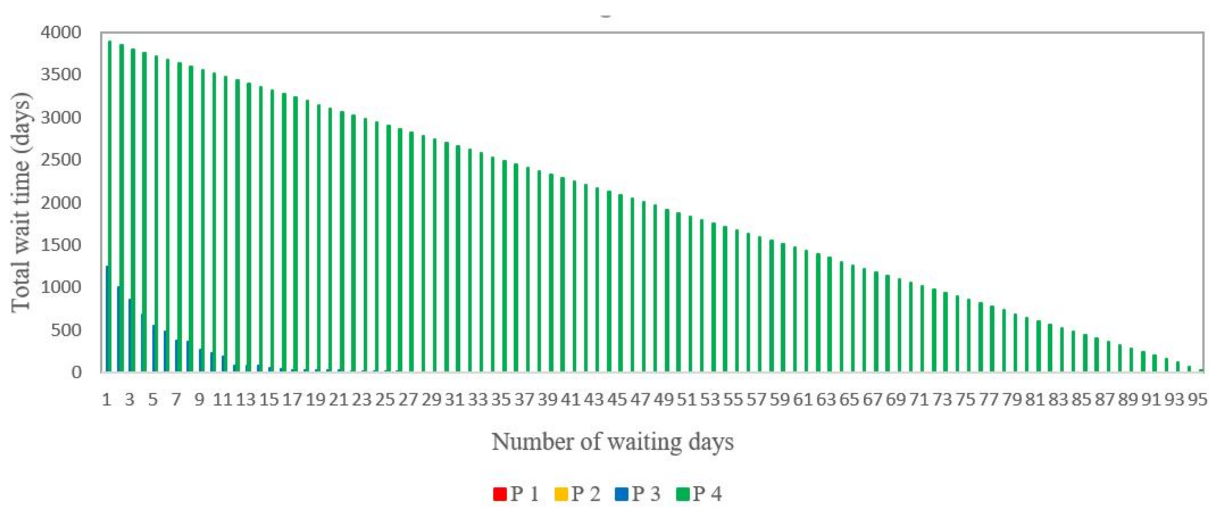


Figure 4.15: Wait time of patients in each priority, $C = 105$, $\Gamma_p = 100$

Figure 4.15 shows wait times of patients who have waited n days to receive the service, without forcing the desired service level of 90%. Comparing Figure 4.15 and Figure 4.8, it is noted that the total wait times of patients priority 4 in strict policy are significantly higher than those in the proposed approach. The analysis shows that by delaying patients of higher priority within their wait time target while meeting the desired service level, lower priority wait times are mitigated.

4.8. Robust versus Deterministic

To determine the effectiveness of robust optimization approach for scheduling problem, we compare robust schedule against deterministic schedule under different scenarios. Section 4.8.1 details the data used and scenarios generated.

4.8.1 Scenarios and datasets

Based on Ontario's hospital data, Jiang (2016) shows that daily customer arrivals follow a normal distribution with mean $\mu = 41.3$ and standard deviation $\sigma = 12.8$. We use the same normal distribution $N(41.3, 12.8)$, to randomly generate the daily number of patients over the planning horizon T. Once the daily number of patient arrivals are estimated, it is proportioned between different priorities under three scenarios.

Based on 74 MRI hospitals in Ontario, Jiang (2016) reports that daily patient arrivals are categorized into 4 priorities with following proportions: where 2% are categorized as priority 1, 7% priority 2, 14% priority 3, and 77% are priority 4. We use these proportions to generate scenario 1. as shown in Table 4.3. To investigate the effect of the different proportion of demand, we generate scenario 2 and 3. In scenario 2, we assume that daily demand is relatively proportioned equally among different priorities. In scenario 1, we note that higher priority patients have lower demand. In contrast to scenario 1, we generate scenario 3 in which patients with high priority level have higher demand.

Scenario	% Patients of priority p			
	1	2	3	4
1	2	7	14	77
2	20	23	25	32
3	77	14	7	2

Table 4.3: Summary of scenarios

A random dataset is generated for each scenario which is used to solve the robust and deterministic models. For each dataset, we estimate uncertainty interval for each patient $p \in P$ as $[\min_{t\in T}\{d_{pt}\}, \max_{t\in T}\{d_{pt}\}]$. The nominal value is then calculated as $\overline{d}_{pt} = \frac{\min_{t\in T}\{d_{pt}\} + \max_{t\in T}\{d_{pt}\}}{2}$. These values are used to solve the robust model, and for the deterministic model, we directly use randomly generated daily demand values. For each scenario, we generate robust and deterministic schedule which are tested over 10 new random instances. We also solve each scenario over four different capacity levels to see the effect of capacity on the difference between the schedule before and after the realization of arrivals. Three metrics are defined: the total average wait time for all patients ($\overline{wt}$), the average wait time for each priority (wt_p), and the service level (α_p). The averages of data in 10 datasets are reported in Tables 4.4, 4.5, 4.6 for scenario 1, scenario 2 and scenario 3, respectively. In these tables, R and D denote the robust and the deterministic model results, respectively. The percentage of improvements are reported as the difference between the robust approach and the deterministic one, $\Delta\%$. It shows how the schedule from the robust approach performs better than the schedule from the deterministic model.

In Table 4.4, we observe that for all capacities, the robust approach outperforms the deterministic model. For example, when capacity is 47 hours, priority 1 patients wait for an average of 0.2 days under the robust schedule, while it is 2.8 days in the deterministic one. Also, the service level of priority 1, which shows the number of patients who received service

Capacity	Model	Average wait time, wt_p					Service level, α_p			
		wt_1	wt_2	wt_3	wt_4	$\overline{wt}$	α_1	α_2	α_3	α_4
47	R	0.2	0.8	1.1	0.5	0.6	96.4%	87.9%	93.0%	100.0%
	D	2.8	3.0	2.7	2.5	2.6	68.0%	72.7%	90.9%	98.0%
	Δ%	-92%	-73%	-59%	-82%	-78%	42%	21%	2%	2%
48	R	0.0	0.3	0.2	0.4	0.4	99.9%	93.9%	99.6%	99.5%
	D	2.5	1.9	1.7	1.6	1.7	73.5%	79.1%	94.7%	99.5%
	Δ%	-100%	-84%	-90%	-75%	-78%	36%	19%	5%	0%
49	R	0.0	0.3	0.2	0.2	0.2	99.8%	93.7%	99.9%	99.6%
	D	2.5	1.9	1.7	1.6	1.7	73.0%	79.6%	94.6%	99.3%
	Δ%	-100%	-83%	-87%	-86%	-86%	37%	18%	6%	0%
50	R	0.0	0.3	0.2	0.1	0.1	100.0%	94.6%	99.9%	99.9%
	D	2.5	1.9	1.7	1.6	1.7	72.8%	79.2%	94.9%	99.3%
	Δ%	-100%	-87%	-88%	-92%	-92%	37%	19%	5%	1%

Table 4.4: Robust versus Deterministic optimization- Scenario 1

Capacity	Model	Average wait time, wt_p					Service level, α_p			
		wt_1	wt_2	wt_3	wt_4	$\overline{wt}$	α_1	α_2	α_3	α_4
48	R	1.1	1.1	0.9	3.4	1.8	78.7%	88.6%	94.0%	94.3%
	D	5.6	5.5	6.4	6.1	5.9	65.0%	70.8%	85.2%	91.2%
	Δ%	-80%	-80%	-87%	-44%	-70%	21%	25%	10%	3%
49	R	0.7	0.4	0.9	2.4	1.2	81.6%	91.6%	93.8%	96.8%
	D	5.6	5.5	6.4	6.0	5.9	64.5%	70.9%	86.8%	91.0%
	Δ%	-87%	-92%	-86%	-60%	-79%	27%	29%	8%	6%
50	R	0.9	0.2	0.4	1.9	0.9	80.3%	94.6%	96.6%	95.5%
	D	5.6	5.5	6.4	6.0	5.9	66.8%	72.2%	85.8%	91.5%
	Δ%	-84%	-96%	-94%	-68%	-84%	20%	31%	13%	4%
51	R	0.8	0.2	0.2	1.2	0.6	81.0%	95.5%	99.6%	93.3%
	D	5.6	5.5	6.4	5.9	5.9	65.9%	71.3%	85.9%	91.7%
	Δ%	-86%	-97%	-97%	-81%	-90%	23%	34%	16%	2%

Table 4.5: Robust versus Deterministic optimization - Scenario 2

within 1 day, is 96.4% in the robust schedule, whereas the deterministic schedule serves 68% of priority 1 on the day of arrival. We also note that as the capacity increases, the difference of average wait time of some priorities increases in favor of the robust approach.

As shown in Table 4.5, the schedule of the robust approach performs better than that of the deterministic model even when demand can take values outside of the uncertainty set. We expect that even if the number of deviating demand points is greater than $\Gamma_p = 100$,

Capacity	Model	Average wait time, wt_p					Service level, α_p			
		wt_1	wt_2	wt_3	wt_4	$\overline{wt}$	α_1	α_2	α_3	α_4
47	R	0.4	2.1	3.0	6.0	1.0	86.3%	82.8%	90.1%	89.6%
	D	2.8	3.1	3.5	5.8	2.9	56.9%	67.6%	91.5%	93.7%
	Δ%	-84%	-35%	-13%	3%	-67%	52%	22%	-2%	-4%
48	R	0.3	2.0	2.9	5.9	0.8	89.8%	82.7%	90.2%	90.0%
	D	2.8	3.1	3.4	5.3	2.9	58.6%	69.6%	90.8%	94.2%
	Δ%	-90%	-37%	-14%	11%	-72%	53%	19%	-1%	-4%
49	R	0.2	1.6	2.3	5.1	0.6	91.7%	84.4%	91.6%	90.9%
	D	2.8	3.1	3.3	4.7	2.9	56.7%	68.8%	90.5%	94.8%
	Δ%	-93%	-48%	-29%	8%	-78%	62%	23%	1%	-4%
50	R	0.1	0.9	1.6	3.9	0.4	93.3%	88.0%	93.9%	92.8%
	D	2.8	3.0	3.1	4.5	2.9	56.3%	68.9%	91.6%	94.8%
	Δ%	-96%	-69%	-49%	-14%	-85%	66%	28%	2%	-2%

Table 4.6: Robust versus Deterministic optimization - Scenario 3

robust performs better with a high probability (Bertsimas and Sim, 2004). As noted, the average wait times of patients under the robust schedule are less than those in the deterministic model, and the service levels are higher in the robust approach, in all cases.

The results for scenario 3 are summarized in Table 4.6 where we note that for higher priority patients, the robust model outperforms the deterministic model results both in terms of average wait time and service level achieved. For example, when capacity is 49, the average wait time and service level of priority 1 patients are 0.2 days and 91.7 %, respectively, while it outperforms deterministic by 93% reduction in average wait time and 62% increase in service level. Furthermore, in case of the average wait time of patients independent of priority levels, robust performs better. When capacity is 50 hours, patients wait for an average of 0.4 days under the robust schedule, whereas they wait for an average of 2.9 days in the deterministic one. Although for lower priority patients, the deterministic schedule performs better than the robust schedule, the difference is not significant.

Chapter 5

Conclusion and Future Research

In this work, we studied the problem of scheduling multi-priority patients over a multi-period horizon while ensuring to meet a desired level of service. Due to demand uncertainty, a deterministic model may not perform well, and as such, we developed a robust optimization approach to making a schedule which is robust against uncertainty in daily patient arrivals. We exploited the notion of the budget of uncertainty to control the level of performance against the level of conservatism. The budget of uncertainty allows a decision maker to adjust the budget of uncertainty based on their preferences for each priority level. The proposed robust method is applied to several datasets, and an extensive sensitivity analysis is performed over important parameters such as capacity, penalty cost, service level, and budget of uncertainty.

The analysis showed that the proposed robust approach outperforms the deterministic method concerning service level and patient wait time. Furthermore, we observed that even if the realized demand occurs outside the defined interval for uncertain demand, RO solutions provide higher service levels and lower wait times compared to those in the deterministic approach. Extensive sensitivity analysis presents key managerial insights with respect to the effect of capacity, penalty cost, and budget of uncertainty on the optimal schedule.

Our work has some limitations that could be addressed in the future research. One possible extension could be to consider different types of MRI scans with various service durations, and also incorporating service time uncertainties into the model. Furthermore, sensitivity analysis can be performed on the length of the planning horizon, various service levels for different priorities, and different wait time targets.

We may also employ a dynamic decision-making approach where a schedule can be updated based on the realized demand. The problem can be formulated as a two-stage robust optimization problem. In two-stage decision making, there are two sets of decisions; first stage decisions and recourse decisions, where recourse decisions are decisions that are made after the realization of an uncertain parameter.

Another possible extension is to apply Pareto robust optimization (PRO). Traditional robust optimization approach may lead to multiple optimal solutions. By utilizing PRO, we may find solutions that dominate RO solutions under non-worst case scenarios while maintaining the performance of RO solution under the worst-case scenario.

References

Astaraky, Davood and Patrick, Jonathan. A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. *European Journal of Operational Research*, 245(1):309–319, 2015.

Ayvaz, Nur and Huh, Woonghee Tim. Allocation of hospital capacity to multiple types of patients. *Journal of Revenue and Pricing Management*, 9(5):386–398, 2010.

Baron, Opher; Milner, Joseph, and Naseraldin, Hussein. Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785, 2011.

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

Ben-Tal, Aharon and Nemirovski, Arkadi. Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13, 1999.

Ben-Tal, Aharon and Nemirovski, Arkadi. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3):411–424, 2000.

Ben-Tal, Aharon and Nemirovski, Arkadi. Selected topics in robust convex optimization. *Mathematical Programming*, 112(1):125–158, 2008.

Ben-Tal, Aharon; El Ghaoui, Laurent, and Nemirovski, Arkadi. *Robust optimization.* Princeton University Press, 2009.

Bertsimas, Dimitris and Sim, Melvyn. Robust discrete optimization and network flows. *Mathematical programming*, 98(1):49–71, 2003.

Bertsimas, Dimitris and Sim, Melvyn. The price of robustness. *Operations research*, 52(1): 35–53, 2004.

Bertsimas, Dimitris and Thiele, Aurélie. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

Bertsimas, Dimitris; Pachamanova, Dessislava, and Sim, Melvyn. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

Bertsimas, Dimitris; Iancu, Dan A, and Parrilo, Pablo A. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

Bertsimas, Dimitris; Brown, David B, and Caramanis, Constantine. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

Bertsimas, Dimitris; Litvinov, Eugene; Sun, Xu Andy; Zhao, Jinye, and Zheng, Tongxin. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013.

Birge, John R and Louveaux, Francois. *Introduction to stochastic programming.* Springer Science & Business Media, 2011.

Bortfeld, Thomas; Chan, Timothy CY; Trofimov, Alexei, and Tsitsiklis, John N. Robust management of motion uncertainty in intensity-modulated radiation therapy. *Operations Research*, 56(6):1461–1473, 2008.

CADTH, . Canadian agency for drugs and technologies in health, 2015. URL https://www.cadth.ca/canadian-medical-imaging-inventory-2015.

CAR, . National maximum wait time access targets for medical imaging (mri and ct), 2013. URL http://www.car.ca/uploads/standards%20guidelines/CAR_National_Maximum_WaitTime_Targets_MRI_and_CT_2013_EN.pdf.

Cardoen, Brecht; Demeulemeester, Erik, and Beliën, Jeroen. Operating room planning and scheduling: A literature review. *European journal of operational research*, 201(3): 921–932, 2010.

Cayirli, Tugba and Veral, Emre. Outpatient scheduling in health care: a review of literature. *Production and operations management*, 12(4):519–549, 2003.

Chakraborty, Santanu; Muthuraman, Kumar, and Lawley, Mark. Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*, 42(5):354–366, 2010.

Chan, Timothy CY; Mahmoudzadeh, Houra, and Purdie, Thomas G. A robust-cvar optimization approach with application to breast cancer therapy. *European Journal of Operational Research*, 238(3):876–885, 2014.

Chen, Daizhuo; Meng, Fanwen; Ang, James; Chu, Singfat; Sim, Melvyn, and Kannapiran, Palvannan. A robust optimization model for managing elective admission in hospital. Technical report, Technical report, National University of Singapore, Singapore, 2011.

Chen, Xin; Sim, Melvyn, and Sun, Peng. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

CIHI, . Wait times for priority procedures in canada, Mar 2016. URL https://secure.cihi.ca/estore/productFamily.htm?locale=en&pf=PFC3108.

CIHI, . Wait times for priority procedures in canada, Apr 2017. URL https://www.cihi.ca/en/wait-times-for-priority-procedures-in-canada-2017.

Denton, Brian and Gupta, Diwakar. A sequential bounding approach for optimal appointment scheduling. *Iie Transactions*, 35(11):1003–1016, 2003.

Dong, Jing; Yom-Tov, Elad, and Yom-Tov, Galit B. The impact of delay announcements on hospital network coordination and waiting times. Technical report, Technical report, Working Paper, 2015.

El Ghaoui, Laurent; Oustry, Francois, and Lebret, Hervé. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

Fabozzi, Frank J; Huang, Dashan, and Zhou, Guofu. Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1):191–220, 2010.

Feldman, Jacob; Liu, Nan; Topaloglu, Huseyin, and Ziya, Serhan. Appointment scheduling under patient preference and no-show behavior. *Operations Research*, 62(4):794–811, 2014.

Gabrel, Virginie; Murat, Cécile, and Thiele, Aurélie. Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483, 2014.

Gerchak, Yigal; Gupta, Diwakar, and Henig, Mordechai. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42(3): 321–334, 1996.

Gocgun, Yasin and Ghate, Archis. Lagrangian relaxation and constraint generation for allocation and advanced scheduling. *Computers & Operations Research*, 39(10):2323–2336, 2012.

Gocgun, Yasin and Puterman, Martin L. Dynamic scheduling with due dates and time windows: an application to chemotherapy patient appointment booking. *Health care management science*, 17(1):60–76, 2014.

Gorissen, Bram L; Yanıkoğlu, İhsan, and den Hertog, Dick. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

Green, Linda V; Savin, Sergei, and Wang, Ben. Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1):11–25, 2006.

Guerriero, Francesca and Guido, Rosita. Operational research in the management of the operating theatre: a survey. *Health care management science*, 14(1):89–114, 2011.

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

Gupta, Diwakar and Wang, Lei. Revenue management for a primary-care clinic in the presence of patient choice. *Operations Research*, 56(3):576–592, 2008.

Hanne, Thomas; Melo, Teresa, and Nickel, Stefan. Bringing robustness to patient flow management through optimized patient transports in hospitals. *Interfaces*, 39(3):241–255, 2009.

Hazır, Öncü; Haouari, Mohamed, and Erel, Erdal. Robust scheduling and robustness measures for the discrete time/cost trade-off problem. *European Journal of Operational Research*, 207(2):633–643, 2010.

Iancu, Dan A and Trichakis, Nikolaos. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2013.

Jiang, Yangzi. Data-driven analytics to support scheduling of multi-priority multi-class patients with wait targets. Master's thesis, 2016. URL http://hdl.handle.net/10012/10615.

Kim, Seung-Chul and Horowitz, Ira. Scheduling hospital services: the efficacy of elective-surgery quotas. *Omega*, 30(5):335–346, 2002.

Kleinrock, Leonard and Finkelstein, Roy P. Time dependent priority queues. *Operations Research*, 15(1):104–116, 1967.

Kolesar, Peter. A markovian model for hospital admission scheduling. *Management Science*, 16(6):B–384, 1970.

Li, Na. Recent advances in accumulating priority queues. 2015.

Lowery, Julie C and Martin, James B. Evaluation of an advance surgical scheduling system. *Journal of medical systems*, 13(1):11–23, 1989.

Luo, Jianzhe; Kulkarni, Vidyadhar G, and Ziya, Serhan. Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management*, 14(4):670–684, 2012.

Magerlein, James M and Martin, James B. Surgical demand scheduling: a review. *Health services research*, 13(4):418, 1978.

Mamani, Hamed; Nassiri, Shima, and Wagner, Michael R. Closed-form solutions for robust inventory management. *Management Science*, 63(5):1625–1643, 2016.

Meng, Fanwen; Qi, Jin; Zhang, Meilin; Ang, James; Chu, Singfat, and Sim, Melvyn. A robust optimization model for managing elective admission in a public hospital. *Operations Research*, 63(6):1452–1467, 2015.

Min, Daiki and Yih, Yuehwern. An elective surgery scheduling problem considering patient priority. *Computers & Operations Research*, 37(6):1091–1099, 2010.

MOHLTC, . Ontario wait times, 2017a. URL http://www.health.gov.on.ca/en/pro/programs/waittimes/surgery/target.aspx.

MOHLTC, . Surgery and diagnostic imaging results, 2017b. URL http://www.ontariowaittimes.com/SurgeryDI/EN/provincialsummary.aspx.

Mondschein, Susana V and Weintraub, Gabriel Y. Appointment policies in service operations: A critical analysis of the economic framework. *Production and Operations Management*, 12(2):266–286, 2003.

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

Parizi, Mahshid Salemi and Ghate, Archis. Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking. *Computers & Operations Research*, 67:90–101, 2016.

Patrick, Jonathan; Puterman, Martin L, and Queyranne, Maurice. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations research*, 56(6):1507–1525, 2008.

Qu, Xiuli; Rardin, Ronald L; Williams, Julie Ann S, and Willis, Deanna R. Matching daily healthcare provider capacity to demand in advanced access scheduling systems. *European Journal of Operational Research*, 183(2):812–826, 2007.

Saure, Antoine; Patrick, Jonathan; Tyldesley, Scott, and Puterman, Martin L. Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research*, 223(2):573–584, 2012.

Soyster, Allen L. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157, 1973.

Stanford, David A; Taylor, Peter, and Ziedins, Ilze. Waiting time distributions in the accumulating priority queue. *Queueing Systems*, 77(3):297–330, 2014.

Truong, Van-Anh. Optimal advance scheduling. *Management Science*, 61(7):1584–1597, 2015.

Wald, Abraham. Generalization of a theorem by v. neumann concerning zero sum two person games. *Annals of Mathematics*, pages 281–286, 1945.

WTA, . Canadians still waiting too long for health care, 2013. URL www.waittimealliance.ca/wp-content/uploads/2014/05/2013_Report_Card.pdf.