OPTIMAL OVERRIDE POLICY FOR CHEMOTHERAPY SCHEDULING TEMPLATE VIA MIXED-INTEGER LINEAR PROGRAMMING

by

Ishtiak Sikder

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Engineering Management

Charlotte

2021

Approved by:

____
Dr. Guanglin Xu

____
Dr. Churlzu Lim

____
Dr. Linquan Bai

____
Dr. Yu-Li Huang

ABSTRACT

ISHTIAK SIKDER. Optimal Override Policy for Chemotherapy Scheduling Template via Mixed-Integer Linear Programming. (Under the direction of DR. GUANGLIN XU)

Owing to treatment complexity in chemotherapy administration, nurses are usually required at the beginning, end, and at certain times during treatment to ensure high-quality infusion. It is, thus, critical for an outpatient chemotherapy unit to design a scheduling template that can effectively match nursing resources with treatment requirements.

The template contains appointment slots of different lengths and thus allows schedulers to place patients into these appointment slots according to the provider's order. As the template is often used over a period of several months, there usually exists a mismatch between the daily patient mix and the fixed structure of the given template.

Hence, override policies must be employed to adjust to demand. However, these policies are often manually performed by schedulers.

A Mixed-Integer Linear Programming (MILP) model has thus been proposed in this thesis to systematically develop optimal override policies in place of the manual process to improve template utilization while maintaining template stability. Numerical experiments based on real-life data from a chemotherapy unit are conducted to demonstrate the effectiveness of the proposed approach.

DEDICATION

This thesis is dedicated to my loving parents, without whose continued motivation and support I never would have made it this far.

ACKNOWLEDGEMENTS

First and foremost, I am deeply indebted to my thesis chair, Dr. Guanglin Xu for seeing the potential in me and including me in his research projects from my very first semester at UNCC. The work that I have had the privilege to do with him and Dr. Yu-Li Huang, assistant professor at Mayo Clinic, forms the basis of this thesis and has taught me a lot in applying operations research theories to solve real-world problems. I am immensely grateful to Dr. Yu-Li Huang as well for being such a kind mentor to me and for guiding me throughout our collaboration.

I would also like to express my deepest gratitude towards Dr. Churlzu Lim and Dr. Linquan Bai for accepting to join my thesis committee. Special thanks to Dr. Churlzu Lim for his kind support throughout the graduate courses I took with him, where I learnt the required theories applied to complete this thesis work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MILP Mixed-Integer Linear Programming

PREFACE

The content of this thesis is based on the research paper of the same title accepted for publication under the Optimization Letters journal, co-authored by me along with my research advisors, Dr. Guanglin Xu from UNCC and Dr. Yu-Li Huang from Mayo Clinic.

CHAPTER 1: INTRODUCTION

Healthcare is a significant industry all over the world, as it protects normal functioning which in turn protects the range of opportunities open to individuals [6]. To provide some context, on average the United States spends more than 2 trillion or 16% of its GDP on Healthcare each year [3]. Common challenges in health care administration often revolves around ensuring efficient resource distribution for providing adequate treatment to patients, which makes it a relevant focus for operations research studies. A common application of OR in health care is to develop scheduling templates for outpatient appointments, with related operations research work going as far back as the early 1950s [1].

An outpatient service usually serves patients on a daily basis based on pre-scheduled appointments. Unlike inpatient services, this mode of treatment allows the patient to come in, receive treatment, and leave within the same day without having to enroll for an extended stay. The advancement in medicare has made outpatient treatment services possible for Chemotherapy administration, which solely relied on inpatient treatment in the past. Due to the nature of Chemotherapy treatment, such an outpatient service centre typically deals with patients requiring varying time lengths of treatment; which can be anything between 30 minutes and 6 hours or more. Optimized solutions abound to propose appointment scheduling templates that match available nursing resources with scheduled patients for safe infusion. But managing the uncertainty of the daily patient influx to utilize the template to its maximum capacity while meeting nursing resource constraints - presents an optimization problem in itself, which has largely remained unsolved.

The scheduling template typically has allotments for different types of patients according to the treatment length required for them. For example, if the working day is divided into 15 minute time slots then an appointment slot for an outpatient requiring a 30 minute chemotherapy infusion will take up (30/15=) 2 time slots on the scheduling template and said patient can be labeled as a 30 minute patient. Same is true for a 60 minute patient or a 120 minute patient and so forth. Having established that, a chemotherapy outpatient service can either have more or less number of a given type of patient on a given day than what is allotted in the scheduling template, i.e (n+2) number of 30 minute patients and (n-1) number of 60 minute patients. It is to be noted that the scheduling template is designed to accommodate necessary nursing resources for scheduled patients. Thus, having less patients of a given type (i.e (n-1) 60 min patients) would mean a wastage of allotted nursing resources; which could possibly serve extra patients of another type (i.e 2 extra 30 minute patients for an influx of (n+2) 30 minute patients) if said extra patients could be accommodated. This can provide a brief outlook to the challenge with the variability of the daily patient influx.

An obvious solution can be to accommodate extra patients into available empty time slots, i.e the two extra 30 minute patients can be accommodated into the empty 60 minute appointment slot for an influx of (n+2) 30 minute patients and (n-1) 60 minute patients. This allotment practice is termed as an "Override" and is often applied manually by chemotherapy outpatient appointment schedulers. But the manual application leaves room for a lot of scheduling errors. The schedulers usually prioritize accommodating as many extra patients into empty slots as possible without regarding nursing resource constraint violations, and even if they wanted to look out for resource constraints it would be too complex to manually override for a scheduling template containing up to 7 or more different types of Chemotherapy outpatients. A systematic approach is thus necessary to solve this problem.

This study introduces a novel optimization scheme to generate optimal override policies that may be applied to assign appointment slots in a fixed template to accommodate patients requiring different treatment lengths. Particularly, a mixed-integer linear programming model for determining the optimal override policies has been proposed and solved by using an off-the-shelf solver. Numerical experiments on real data from a chemotherapy unit are conducted to demonstrate the effectiveness of the proposed approach. This is the first study to utilize mixed-integer linear programming techniques to derive optimal policies in the context of chemotherapy appointment scheduling.

The remaining chapters are organized as follows. Chapter 2 reflects on relevant literature regarding appointment scheduling and process variability in Chemotherapy administration. Chapter 3 discusses the chemotherapy appointment scheduling template and the override policies, and propose an MILP model to optimally apply the override policies. Chapter 4 discusses the design of the numerical experiments to demonstrate the effectiveness of the proposed approach over real-life data from a chemotherapy unit. Finally, concluding remarks and future extensions for optimization-based override policy problems are discussed in Chapter 5.

CHAPTER 2: LITERATURE REVIEW

The safe administration of chemotherapy is paramount and, therefore, the optimal management of workflows in the chemotherapy units has become a major focus for any practical setting [13]. In the past decades, chemotherapy administration has gradually shifted from the inpatient setting to the outpatient setting due to the advanced development in medical delivery methods, drug and prescription innovation, and side effect management [23]. In general, outpatient service involves patients coming into the hospital to obtain essential health care and leaving within the day after receiving the service. The standard method of scheduling outpatient appointments is to apply a policy that accounts for appointment block size and intervals; see e.g., [4].

Appointment scheduling has been an extensively studied research area in operations research and management science [1], [10], [19], [20], since it has applications in a wide range of fields such as outpatient scheduling [9], [12], [16], [18], surgery planning [7], call-center staffing [11], and cloud computing server operations [22]. It often aims to efficiently allocate scarce resources to satisfy requirements against some physical and/or economic constraints. For instance, the primary operational objective of many appointment systems is to design scheduling templates for appointments to optimize the overall benefit or costs of the system. Generally, there are two typical types of appointment systems: single- and multi-server systems; see [15], [24].

One of the main challenges to develop a flexible appointment scheduling system, regardless of the type, is the process variability, of which there are several. First, although the allotted time for a patient in the scheduling template is fixed, the actual appointment length is typically unknown. For instance, the treatment length of a patient may depend on their health state, which has to be observed on the fly. There is also uncertainty due to the requirement of urgent appointments, walk-ins, and the occurrence of patients not showing for appointments. Random no-shows cause poor resource utilization and unanticipated loss of revenue for health care providers, which may, in turn, compromise service quality [2]. Furthermore, the varying patient mix also causes dramatic modeling challenges for the scheduling system, especially in chemotherapy treatment facilities. In fact, tackling this type of variability is the main focus of this research. Stochastic programming [17] and simulation are generally utilized as solution methods to solve scheduling problems with uncertainties; see e.g., [8], [21], and [5].

In practice, treatment plan schedules for chemotherapy may vary depending on the type of cancer, the associated treatment regimen and its goal, as well as the patients' state of health. An oncologist decides the choice of a particular regimen, but modifications in drug dosage and schedules are often necessary due to the variability in the health status of a patient. Therefore, a scheduling template must be determined to consist of appointment slots that can accommodate a patient mix requiring various treatment lengths. In other words, the appointment slots in the template should be of different lengths based on the patient mix with various treatment requirements. However, a fixed scheduling template may not be able to accommodate varying patient mix on a daily basis. To this end, override policies should be employed to modify the existing appointment slots to accommodate patients. In particular, multiple appointment slots can be combined into one to serve a patient requiring a longer treatment. A slot may also be broken into multiples to serve those patients requiring shorter treatments; as well as longer appointment slots being divided into shorter ones. In practice, it is challenging to implement these override policies manually.

In this research, an MILP model has been developed to derive optimal override policies. More specifically, assuming that a fixed template and a patient mix are given, the MILP model selects existing appointment slots to treat patients via pre-determined override policies (see Section 2.0.1). The model can provide schedulers with flexibility in the preference of the override policies. The MILP model has been verified over a real-life data set collected from a chemotherapy unit at the Mayo Clinic. The result from numerical experiments illustrates the effectiveness of the proposed approach.

It is to be noted that although many literature articles also study methods to address mismatch between the template and patient mix, they focus mainly on designing flexible templates. This research marks the first initiative to develop optimization methods to modify the pre-determined template to accommodate varying patient mix on the daily basis.

CHAPTER 3: PROBLEM FORMULATION

This chapter discusses the override policies that are used to modify appointment time to handle potential mismatch between the existing template with appointment of fixed lengths and the daily patient mix requiring different treatment lengths. To search for the optimal override policies, an MILP model is proposed that can be solved by using off-the-shelf solvers such as Gurobi or CPLEX. In what follows, it is assumed that a given scheduling template is deployed at the chemotherapy unit and that the patient mix of different types are available when the scheduler sets the appointments.

Chemotherapy units often utilize a template over a period of several months or even years. To ensure that a template accounts for nursing resource needs during patient treatments, a fixed template is often proposed. The template incorporates certain numbers of each patient type. Here patient types are categorized based on their required treatment lengths. The treatment length of each patient type is usually designed to have enough buffer to account for activities in addition to administering chemo drugs. This buffer also includes the time to clean the chemo space. No additional time is needed between appointments, even though the distribution of the different types of incoming patients varies on a daily basis. This phenomenon leads to a mismatch between the appointment slots designed by the template and patients who require different treatment lengths. To handle this mismatch, the scheduler often manually follows override policies to modify the appointment template in order to accommodate patients in the current practice at Mayo Clinic [14]. This manual process is time-consuming and often increases workload for nurses if it is not done properly. Hence, a systematic approach with objectives that attempt to align with the scheduling template is beneficial to maintaining template efficiency.

3.1 Override policies for chemotherapy scheduling template

In this section, three override policies are proposed that are applied to guide the modification of the template in the presence of mismatches between the patient mix and the designed appointment slots. By the clinic requirement, the override should incorporate the following three policies:

- *Policy 1*: A patient is placed in a longer appointment when the designated appointment slot is no longer available. For example, a patient requiring a 2-hour appointment slot can be scheduled in a 3-hour appointment slot or any appointment slots longer than two hours.
- *Policy 2*: Two subsequent appointment slots are combined to create an appointment slot that is equal to or greater than the required treatment length. For example, if a chair/ bed has a 1-hour appointment slot from 7:45 AM to 8:45 AM and a 2-hour appointment slot from 8:45 AM to 10:45 AM, then these can be combined to serve a patient requiring a 3-hour appointment slot from 7:45 AM to 10:45 AM.
- *Policy 3*: A longer appointment slot is broken into two shorter appointment slots. For example, a 6-hour appointment slot can be used for two 3-hour appointment slots or a 2-hour appointment slot and a 4-hour appointment slot. It can also be used for two 2-hour appointment slots or a 2-hour appointment slot and a 3-hour appointment slot. It is not preferred to break a longer appointment slot into more than two shorter appointment slots as the nursing resource is not

fully considered. For a 6-hour appointment slot to be used for assigning two 3-hour appointment slots, the nursing resource is not planned at the end of the first 3-hour patient appointment slot and the beginning of the second 3-hour patient appointment slot. In this example, resource constraint violations are increased by two.

The current manual override process at Mayo Clinic is highly driven by patient preferences. Schedulers sometimes manipulate the template and deviate from the override policies by making allocations to fit as many patients as possible in a single template block; even if there are other times available in a day. Thus, policy 3 is often used to accommodate patients, especially those who require a shorter treatment. For override simplicity, schedulers prefer policy 1 over policy 3 and policy 3 over policy 2 since policy 2 requires finding the right appointment slots to combine, which could be time-consuming. In the current state, schedulers are not restricted to deviate from the override policies. For example, schedulers may break an appointment slot to treat three patients, which is a violation of policy 3. Schedulers may also combine three appointment slots to treat a patient, which is a violation of policy 2. Hence, the need for a systematic approach is paramount.

3.2 Mixed-integer linear programming formulation

In this section, an MILP model has been developed to optimize the use of the override policy. More specifically, given a particular mix of patient arrival, the MILP model searches for the optimal patient assignments.

Overall, variables and constraints have been defined to characterize the set of feasible assignment of the appointment slots to the patient mix requiring different treatment lengths. The objective function is defined to prioritize the appointment slot assignment over the override policies.

To simplify the formulation, two necessary index sets have been defined which are specified for ease of exposition. However, the sets can be specified by different requirements.

Sets:

- Set of 15 minute time intervals throughout a working day $\mathcal{T} = \{1, 2, \ldots, 40\}$.
- Set of different patient types $\mathcal{I} = \{30, 60, 120, 180, 240, 300, 360\}$,where the elements in $\mathcal{I}$ denote patient treatment lengths, e.g., 30 indicates that the treatment of this type takes 30 minutes.

Parameters:

- p_i: Number of type i patients for all $i \in \mathcal{I}$.
- $c_{i,t}$: Number of appointments for type i patients that start at time t for all $i \in \mathcal{I}$ and for all $t \in \mathcal{T}$. It is to be noted that the parameters $c_{i,t}$ are determined by the given scheduling template.
- d: The length of the appointment slot. In the thesis, $d = 15$ has been considered.
- λ_k $k = 1, \ldots, 5$: Weight parameters that are used in the objective function.

Variables:

Now the following binary variables are defined to represent an assignment where a type i' appointment slot is used to treat a type i patient ($i \leq i'$). First, a set $[c_{i,t}] := \{1, \ldots, c_{i,t}\}$ is defined to contain all type i appointments that start from time t. Here type i appointments are denoted as the appointments that are planned to treat type i patients in the designed template.

- $x_{i,i',j',t}$: Binary variable, equals 1 if the j'th appointment slot for a type i' patient at time t is used to accommodate a type i patient $(i \leq i')$; and 0 otherwise.

The following binary variables are then defined to represent an assignment where an appointment slot for type i' and an appointment slot for type i'' are combined to treat a patient of type i $(i \leq i' + i'')$.

- $y_{i,i',i'',j',j'',t}$: Binary variable, equals 1 if the j'th appointment slot for a type i' patient at time t and the j''th appointment slot for a type i'' patient at time $t + i'/d$ are combined to treat a type i patient $(i \leq i' + i'')$; and 0 otherwise.

- $z_{i,i',j',t}$: Binary variable, equals 1 if the j'th appointment slot for a patient of type i' at time t is used to start the treatment of a type i patient $(i > i')$; and 0 otherwise.

- $w_{i,i',j',t}$: Binary variable, equals 1 if the j'th appointment slot for a patient of type i' at time t is used to complete the treatment of a type i patient $(i > i')$; and 0 otherwise.

Next, the following binary variables are defined to represent an assignment where an appointment slot of type i' is split into two smaller appointment slots of types i and i'' which can be used to treat two patients.

- $v_{i,i'',i',t,j'}$: Binary variable, equals 1 if the j'th appointment slot for a patient of type i' at time t is divided into two blocks to serve a type i patient and a type i'' patient where $(i + i'' \leq i')$; and 0 otherwise.

Finally, the following auxiliary variables are defined to help model necessary equality constraints.

- $u_{i,t}$: Integer variable, represents the number of type i patients who can be treated at time t.

- $u^0_{i,t}$: Integer variable, represents the number of type i patients who are treated by using overrides at time t.

- q_i: Integer variable, represents the number of type i patients who have not been served.

With the sets, parameters, and variables defined above, we are ready to outline the following constraints of the MILP model.

Constraints:

The following constraints enforce that each appointment slot in the template can be used for a maximum of one task (e.g., equal assignment, larger assignment, combination, and breaking). More specifically, for all $i' \in \mathcal{I}$, for all $t \in \mathcal{T}$, and for all $j \in [c_{i',t}] := \{1, \ldots, c_{i',t}\}$, we have

$$\sum_{i:i\leq i'} x_{i,i',j',t} + \sum_{i:i>i'} z_{i,i',j',t} + \sum_{i:i>i'} w_{i,i',j',t} + \sum_{(i,i'')\in\mathcal{E}} \frac{1}{2} v_{i,i'',i',t,j'} \leq 1 \tag{3.1}$$

where $\sum\limits_{i:i\leq i'}$ denotes the shorthand of notation $\sum\limits_{i\in\mathcal{I}:i\leq i'}$, and $\mathcal{E}$ is defined as follows:

$$\mathcal{E} := \Big\{(i, i'') : i < i', i'' < i', i + i'' \leq i'\Big\}.$$

Note that the set $\mathcal{E}$ is defined to filter out unnecessary variables. For instance, breaking assignments $v_{i,i'',i',j,t'}$ can never happen for the corresponding indices in $\overline{\mathcal{E}}$, which is the complement set of $\mathcal{E}$.

The following constraints are used to ensure the policy to combine two appointments with one of type j and the other of type j' to treat a patient.

$$y_{i,i',i'',j',j'',t} \leq z_{i,i',j',t} \quad \forall\, (i,i',i'',j',j'',t) \in \mathcal{E}_y \tag{3.2}$$

$$y_{i,i',i'',j',j'',t} \leq w_{i,i'',j'',t+i'} \quad \forall\, (i,i',i'',j',j'',t) \in \mathcal{E}_y \tag{3.3}$$

$$y_{i,i',i'',j',j'',t} \geq z_{i,i',j',t} + w_{i,i'',j'',t+i'} - 1 \quad \forall\, (i,i',i'',j',j'',t) \in \mathcal{E}_y \tag{3.4}$$

$$\sum_{i'\in\mathcal{I}} \sum_{j'\in[c_{i',t-i'/d}]} y_{i,i',i'',j',j'',t-i'/d} \leq 1 \quad \forall\, (i,i'',j'',t) \in \mathcal{E}_w \tag{3.5}$$

$$\sum_{i''\in\mathcal{I}} \sum_{j''\in[c_{i'',t}]} y_{i,i',i'',j',j'',t} \leq 1 \quad \forall\, (i,i',j',t) \in \mathcal{E}_z \tag{3.6}$$

where $\mathcal{E}_y, \mathcal{E}_w$, and $\mathcal{E}_z$ are respectively defined as

$$\mathcal{E}_y := \left\{ (i,i',i'',j',j'',t) : \begin{array}{l} i \in \mathcal{I},\ i' \in \mathcal{I},\ i'' \in \mathcal{I} \\ j' \in [c_{i',t}],\ j'' \in [c_{i'',t+i'/d}] \\ i' < i,\ i'' < i,\ i' + i'' \geq i \\ t + (i' + i'')/d \in \mathcal{T} \end{array} \right\},$$

$$\mathcal{E}_w := \left\{ (i,i'',j'',t) : \begin{array}{l} i \in \mathcal{I},\ i'' \in \mathcal{I},\ i > i'' \\ t \in \mathcal{T},\ j'' \in [c_{i'',t}] \end{array} \right\},$$

$$\mathcal{E}_z := \left\{ (i,i',t,j') : \begin{array}{l} i \in \mathcal{I},\ i' \in \mathcal{I},\ i > i' \\ t \in \mathcal{T},\ j' \in [c_{i',t}] \end{array} \right\}.$$

The following constraints ensure the implementation of the breaking policy where a larger appointment slot is used to treat two patients. For all $i \in \mathcal{I}$, for all $t \in T$, and for all $j' \in [c_{i',t}]$, we have

$$\sum_{i' \in \mathcal{I}} \sum_{i'' \in \mathcal{I}} v_{i,i'',i',t,j'} \leq 1. \tag{3.7}$$

Next, we define the following equations which will help establish the objective function. For all $i \in \mathcal{I}$ and for all $t \in \mathcal{T}$, we define

$$\begin{aligned} u_{it} = & \sum_{i':i' \geq i} \sum_{j' \in [c_{i,t}]} x_{i,i',j',t} + \sum_{(i',i'') \in \mathcal{E}'} \sum_{j' \in [c_{i',t}]} \sum_{j'' \in [c_{i'',t}]} y_{i,i',i'',j',j'',t} \\ & + \sum_{(i,i'') \in \mathcal{E}} \sum_{j \in [c_{i,t}]} v_{i,i'',i',t,j'}, \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} u^0_{it} = & \sum_{i':i' > i} \sum_{j' \in [c_{i,t}]} x_{i,i',j',t} + \sum_{(i',i'') \in \mathcal{E}'} \sum_{j' \in [c_{i',t}]} \sum_{j'' \in [c_{i'',t}]} y_{i,i',i'',j',j'',t} \\ & + \sum_{(i,i'') \in \mathcal{E}} \sum_{j \in [c_{i,t}]} v_{i,i'',i',t,j'}, \end{aligned} \tag{3.9}$$

For all $i \in \mathcal{I}$, we define

$$q_i = (p_i - \sum_{t \in \mathcal{T}} u_{i,t})_+. \tag{3.10}$$

Finally, we formulate the mixed-integer linear programming model as follows:

$$\begin{aligned} \min \quad & \lambda_1 \sum_{i \in \mathcal{I}} q_i + \lambda_2 \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} u^0_{i,t} + \lambda_3 \sum_{i \in \mathcal{I}} \sum_{i':i' > i} \sum_{j' \in [c_{i,t}]} \sum_{t \in \mathcal{T}} x_{i,i',j',t} \\ & + \lambda_4 \sum_{(i,i',i'',j',j'',t) \in \mathcal{E}_y} y_{i,i',i'',j',j'',t} + \lambda_5 \sum_{(i,i',i'',t,j) \in \mathcal{E}_v} v_{i,i'',i',t,j'} \end{aligned} \tag{3.11}$$

s.t. eqns (3.1) to (3.10)

Two remarks are made as follows. First, the objective function is used to comprehensively balance the number of unassigned patients and the usage of different policies. Second, the parameters λ_i are used as the priorities of the appointment slot assignments. In particular, λ_1 prioritizes the accommodation of maximum possible number of patients, through either equal assignments or overrides; λ_2 prioritizes equal assignments, which assign appointment slots to patients who require the same treatment length; λ_3 prioritizes patient accommodation through overrides by policy 1; λ_4 prioritizes patient accommodation through overrides by policy 2; and finally λ_5 prioritizes patient accommodation through overrides by policy 3.

CHAPTER 4: NUMERICAL EXPERIMENTS

In this chapter, the MILP model is validated by using real-life data collected from a chemotherapy unit at the Mayo Clinic. All the experiments were implemented in Python 3.7 using Gurobi 8.0 as the MILP solver and were performed on a Macintosh OS X Yosemite system with a quad-core 3.20GHz Intel Core i5 CPU and 8 GB RAM.

4.1 Experimental setup

The experiments were conducted on an actual patient scheduling data set collected from the Andreas Cancer Center, which is situated in Mankato, Minnesota and operated under the Mayo Clinic Health System. Particularly, the data set includes clinical visits from seven different types of patients, which are categorized by the treatment lengths: 30, 60, 120, 180, 240, 300, and 360 minutes. Twenty-two working-day observations were selected in the data set; see Table 4.1. The chemotherapy unit utilizes a fixed template, which is visualized in the left panel in Figure 4.1. The template adopts 14 chairs/beds to accommodate chemotherapy patients on each working day from 7:00 AM to 5:00 PM. Particularly, the template has a total capacity of 61 appointment slots; where 23, 8, 10, 11, 6, 2 and 1 slot(s) have lengths of 30, 60, 120, 180, 240, 300, and 360 minutes respectively.

As shown in Table 4.1, the daily patient mix significantly varies (see, specifically, columns corresponding to *patient number*). Therefore, the scheduler must adopt override policies to accommodate patients' requirements while ensuring their treatment quality. Although the schedulers at Andreas Cancer Center were given instructional policies on how to override template appointment slots [14], it was fairly challenging to follow these policies manually to ensure optimal patient assignments so as to maximize the utilization of the appointment slots. Therefore, the proposed MILP model will be compared with the manual implementations at the Center. Five priority parameters of the model were defined as follows $\lambda_1 = 50, \lambda_2 = 4, \lambda_3 = 1, \lambda_4 = 2, \lambda_5 = 3$ to control optimal patient assignments over the fixed template. Under this setting, the model gives the highest priority ($\lambda_1 = 50$) to allocate appointment slots to treat patients. Moreover, when there is no equal assignment for a patient, the model would have the following preference orders: the larger assignment, the combining override policy, and the breaking override policy ($\lambda_3 = 1 < \lambda_4 = 2 < \lambda_5 = 3$). The experience of using these policies varies among schedulers. The consensus in terms of how long it takes to perform each policy is approximately one minute for policy 1, five minutes for policy 2, and three minutes for policy 3. In addition, to ensure the template's intention of conserving the nurses' availability for the treatment of incoming patients, the impact of breaking an appointment slot (policy 3) on the nurse time is also measured for comparison. Each time an appointment slot is broken to accommodate two appointments, increases two extra appointment slots ($\sim$ 30 minutes) which require nurse resource.

4.2 Solution analysis

In this section, the experimental results are analyzed. Out of the 22 working days, the proposed optimal assignments from the MILP model can completely accommodate the patient influx for 20 working days except for day 13 and day 21. For both of these days, the total patient volume was 62 which translated to 6,840 and 6,780 in total minutes. The volumes exceed the maximum possible number of 61 available appointment slots designed into the template. The total number of required minutes also exceed the maximum number of minutes available, which is 6,750 minutes. In summary, the average daily patient volume was 52.1 with a standard deviation of 6.1. The 22-day data and patient assignments under policy 1, 2, and 3 (from current manual operation and MILP model) are outlined in table 3.1.

Table 3.1 also summarizes the override policy usage for these days. On average, policy 1 has been suggested 2.1 times under the optimal method compared to 3.1 times under the manual method; which shows a 35% reduction in corresponding usage. Policy 2 has been performed 1.2 times on average under the optimal method compared to 2.4 times on average under the manual method, which shows an average reduction of 48% in corresponding overrides. And lastly, policy 3 has been suggested 1.2 times on average under the optimal method compared to 7.9 times on average under the manual method. Policy 3 is the most undesirable override and the optimal method has been able to reduce it by 85%. In overall, the optimal method outperforms the current manual process by 67%, where the override usages have been reduced from 13.4 times to 4.5 times on average. Overall, the results indicate that the proposed MILP model can save 30-minute override time for the schedulers, and 3 work hours for nurses on a regular day.

Figure 3.1 demonstrates via day 16 on how manual (middle panel) and optimal (right panel) methods work. The manual method results in 3 overrides under policy 1, one under policy 2 and 12 under policy 3. For example, a 1-hour appointment is scheduled at 9:30 am in a 3-hour appointment slot starting at 9:15 am on chair 14 ; which meets the criteria for policy 1. On chair 13, the 5-hour appointment slot starting at 9:00 am is combined with a part of 2-hour slot starting at 2:00 pm to be a 6-hour appointment under policy 2. As for policy 3, the 3-hour appointment slot starting at 8:30 am on chair 9 has been broken into a 2-hour (8:30 am) and a 1-hour (10:30 am) appointment. The most undesirable manual override case is found on chair 5 where a 3-hour appointment slot starting at 10:00 am has been broken into three appointments (two 30-minute and one 1-hour). This case often occurs when schedulers attempt to fit patients according to the patients' preference. On the other hand, the optimal method results in 3 overrides under policy 1, one under policy 2, and six under policy 3. This optimal method systematically caps the freedom of breaking an appointment slot while suggests other options for schedulers in order to honor the template design for protecting nurse's time. The average computational time to run the optimization model is approximately 10 seconds over the 22-day data set.

4.3 Sensitivity analysis on the parameters

In this section, sensitivity analysis is conducted on the parameters that are employed to prioritize the override policies in the objective function. In addition to the basic setting for λ_i in Section 4.1, the following settings are also considered.

The sensitivity of λ_1 is investigated first. In the experiment, $\lambda_2 = 4, \lambda_3 = \lambda_4 = \lambda_5 = 1$ are considered while value of λ_1 is varied in the set of $\{10, 30, 50\}$. It is to be noted that a higher value of λ enforces more penalty for not accommodating a patient. As depicted in Figure 4.2(a), the percentage of patients not accommodated decreases as the value of λ_1 increases. Next, the sensitivity of λ_3 is investigated. In the experiment, λ_3 is varied in the set $\{1, 5, 10\}$ and $\lambda_1 = 50, \lambda_2 = 4, \lambda_4 = \lambda_5 = 1$ are maintained. To be noted that a higher value of λ_3 enforces more penalty on applying policy 1. As depicted in Figure 4.2(b), the average number of overrides by policy 1 decreases as the value of λ_3 increases. It can be also observed that the number of overrides by policy 2 remains fairly constant while the number of overrides by policy 3 slightly increases. These results are consistent with the expectation as increasing λ_1 enforces using policies 2 and 3 over policy 1. Sensitivity analysis on λ_4 is then conducted. $\lambda_4 \in \{1, 5, 10\}$ and $\lambda_1 = 50, \lambda_2 = 4, \lambda_3 = \lambda_5 = 1$ are maintained. To be noted

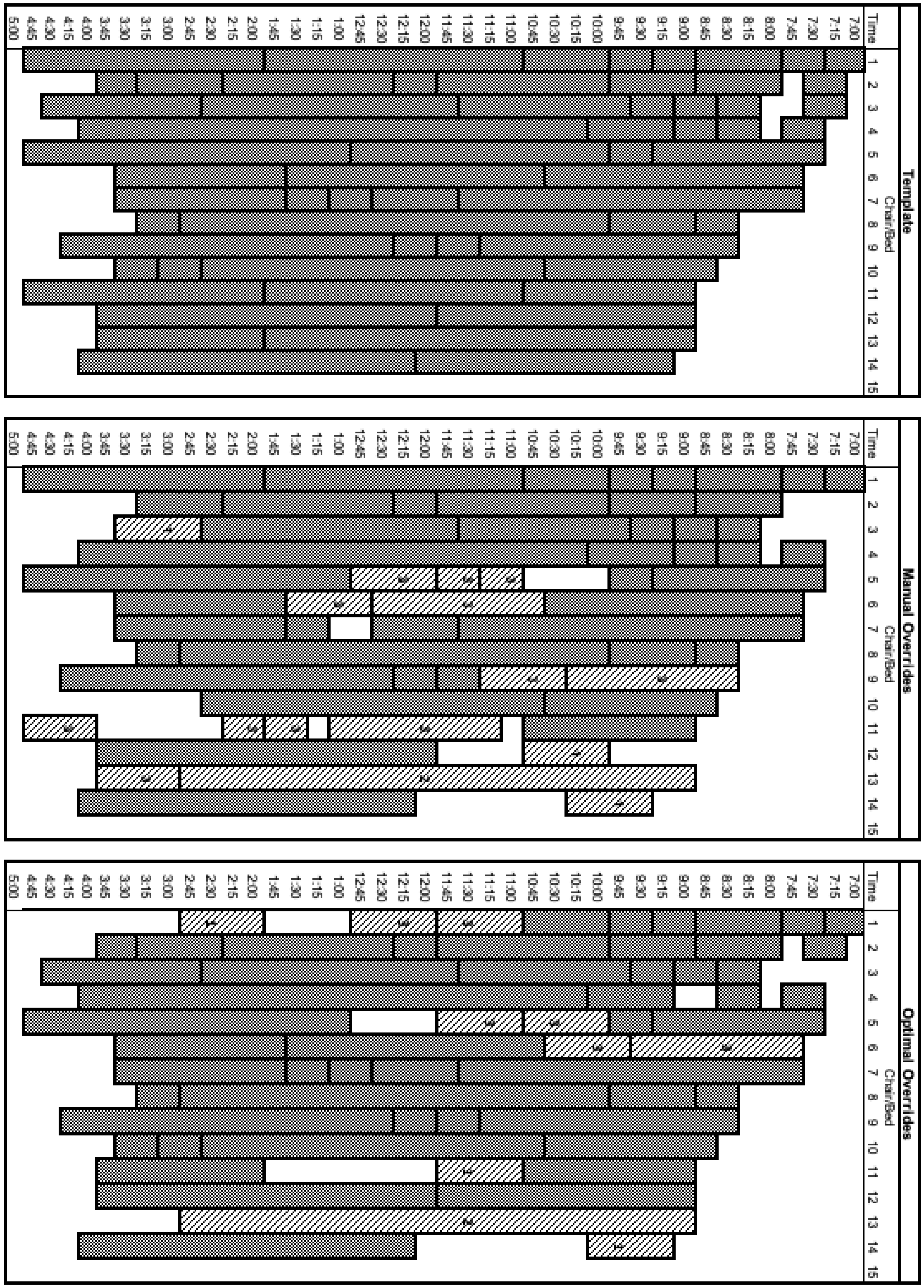

Figure 4.1: The template which is deployed at the chemotherapy unit (left) manual override (middle) and optimal override (right) for day 16.

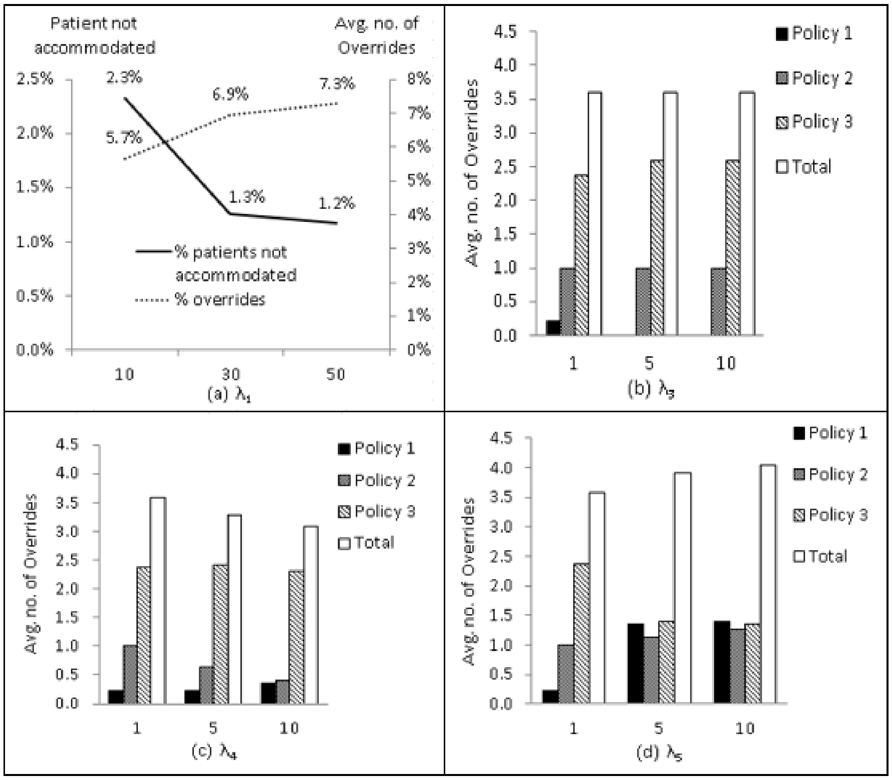

Figure 4.2: Effect of λ_1 on overrides (a), effect of λ_3 on overrides (b), effect of λ_4 on overrides (c), effect of λ_5 on overrides (d)

that a higher value of λ_4 enforces more penalty on applying policy 2. As depicted in Figure 4.2(c), the average number of overrides by policy 2 decreases as the value of λ_4 increases which enforce policies 1 and 3 to accommodate more patients and the average number of overrides by all policies decreases as λ_4 increases. Finally, sensitivity analysis on λ_5 is conducted. $\lambda_5 \in \{1, 5, 10\}$ and $\lambda_1 = 50, \lambda_2 = 4, \lambda_3 = \lambda_4 = 1$ are maintained in the experiment. To be noted that a higher value of λ_5 enforces more penalty on applying policy 3. As depicted in Figure 4.2(d), the average number of overrides by policy 3 decreases as λ_5 increases. It can also be observed that the average number of overrides by policies 1 and 2 increase slightly and that the average number of overrides by all policies increases as λ_5 increases. Furthermore, the total number of overrides is increasing as well.

Table 4.1: 22 days scheduling data from Andreas Cancer Centre at the Mayo Clinic and override results from manual operations and the MILP model.

	Number of patients								Total 2*minutes	Manual policy				Optimal policy			
Day	30	60	120	180	240	300	360	Total		1	2	3	Total	1	2	3	Total
1	15	9	8	5	8	2	0	47	5,370	1	2	7	10	2	1	0	3
2	21	8	5	10	7	2	0	53	5,790	4	1	8	13	1	0	0	1
3	17	10	7	7	9	2	0	52	5,970	3	5	8	16	3	2	0	5
4	15	10	9	5	5	0	2	46	4,950	1	3	6	10	2	1	0	3
5	14	8	10	8	5	3	0	48	5,640	6	3	8	17	1	0	0	1
6	15	13	8	8	6	4	1	55	6,630	4	3	8	15	0	4	4	8
7	20	11	8	7	3	4	0	53	5,400	4	3	8	15	4	1	0	5
8	14	7	8	6	2	3	1	41	4,620	3	1	5	9	0	1	0	1
9	17	13	6	5	3	1	1	46	4,290	3	1	10	14	5	0	0	5
10	18	11	7	6	3	3	2	50	5,460	3	2	7	12	3	2	0	5
11	23	12	9	4	5	0	2	55	5,130	6	3	7	16	4	1	0	5
12	12	8	7	7	6	1	2	43	5,400	5	2	4	11	0	1	0	1
13	24	10	10	7	7	1	3	62	6,840	0	3	13	16	2	1	2	5
14	24	10	12	10	4	1	2	63	6,540	1	1	11	13	0	1	6	7
15	20	9	12	6	6	2	1	56	6,060	4	2	7	13	3	0	0	3
16	21	16	11	4	6	1	2	61	6,090	3	1	12	16	3	1	6	10
17	17	5	7	8	7	1	2	47	5,790	2	1	5	8	1	1	0	2
18	14	9	8	7	5	2	1	46	5,340	2	2	7	11	1	0	0	1
19	14	9	13	3	7	5	1	52	6,600	3	5	7	15	0	8	8	16
20	20	11	6	8	9	1	0	55	5,880	4	4	8	16	5	1	0	6
21	24	8	11	10	5	3	1	62	6,780	3	1	12	16	2	0	0	2
22	20	8	11	5	8	0	1	53	5,580	4	3	6	13	3	0	0	3
Template	23	8	10	11	6	2	1	61	6,750								
									Average number of overrides	3.1	2.4	7.9	13.4	2.1	1.2	1.2	4.5

CHAPTER 5: CONCLUSION AND FUTURE PLANS

In this thesis, override policies are introduced for using a fixed chemotherapy scheduling template to accommodate the varying daily patient mix. In particular, a mixed-integer linear programming model is proposed to determine the optimal override policies, which may significantly mitigate schedulers' work by reducing operating time to find the best appointment slot assignments to patients. Numerical experiments were conducted to demonstrate the effectiveness of the proposed approach. Experimental results also indicated the efficiency of the computation. Furthermore, sensitivity analysis was conducted to demonstrate the flexibility of the proposed approach. Schedulers may tune the parameters to accommodate different preferences for using the model.

It is to be noted that this research considered an existing scheduling template being used at a Chemotherapy Outpatient service and proposed a systematic way for applying optimal overrides. Said template was built from an optimization approach where available nurse distribution was considered as resource constraints, with minimizing nursing resource violations being the primary objective. As already discussed, the variability of the daily patient influx proposes a great challenge in this approach and create nursing resource violations while performing overrides to use this fixed template at maximum possible capacity. The proposed MILP model can thus be employed to evaluate the performance of a given scheduling template and can be incorporated into a high-level optimization problem to determine an optimal scheduling template.

While performing manual overrides, the schedulers are often resorted to apply override policy 3 when patients prefer to be scheduled as early as possible in the day. In other words, they populate the empty available slots early in the day to accommodate as many extra patients as they can, which leads to nursing resource violations. Our proposed MILP model, while efficient in applying minimal overrides to accommodate maximum number of patients, is still limited while taking the patient's appointment scheduling preferences into account. In other words, the model can not override a patient assignment as early as possible in the day to minimize their wait time in queue if said override violates the established override policies. This obviously leaves room for further updates.

As such, an extension of this research can be to build a dynamic model that takes a training data of daily patient influxes and cross checks with available nurse distribution to propose a template with minimal nursing resource violations against applied overrides to reach maximum possible capacity. A two-stage stochastic optimization model can serve this approach well. Constraints can also be introduced to take patient's scheduling preferences into account while also maintaining the established override policies. Thus, the model can produce templates which are already designed to incur minimal resource violations based on provided training data, before being applied against actual patient influx over an extended period of time when overrides will be applied to reach maximum possible capacity and according to schduling preferences of incoming patients.

REFERENCES

[1] Bailey, N.T.: A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. Journal of the Royal Statistical Society. Series B (Methodological) pp. 185–199 (1952)

[2] Berg, B.P., Denton, B.T., Erdogan, S.A., Rohleder, T., Huschka, T.: Optimal booking and scheduling in outpatient procedure centers. Computers & Operations Research **50**, 24–37 (2014)

[3] Catlin, A., Cowan, C., Hartman, M., Heffler, S.: National health spending in 2006: a year of change for prescription drugs. Health Affairs **27**(1), 14–29 (2008)

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

[5] Cayirli, T., Veral, E., Rosen, H.: Designing appointment scheduling systems for ambulatory care services. Health care management science **9**(1), 47–58 (2006)

[6] Daniels, N.: Justice, health, and healthcare. American Journal of Bioethics **1**(2), 2–16 (2001)

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

[8] Erdogan, S.A., Denton, B.: Dynamic appointment scheduling of a stochastic server with uncertain demand. INFORMS Journal on Computing **25**(1), 116–132 (2013)

[9] Fetter, R.B., Thompson, J.D.: Patients' waiting time and doctors' idle time in the outpatient setting. Health services research **1**(1), 66 (1966)

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

[11] Gurvich, I., Luedtke, J., Tezcan, T.: Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. Management Science **56**(7), 1093–1115 (2010)

[12] Ho, C.J., Lau, H.S.: Minimizing total cost in scheduling outpatient appointments. Management science **38**(12), 1750–1764 (1992)

[13] Huang, Y.L., Bach, S.M., Looker, S.A.: Chemotherapy scheduling template development using an optimization approach. International journal of health care quality assurance (2019)

[14] Huang, Y.L., Bryce, A.H., Culbertson, T., Connor, S.L., Looker, S.A., Altman, K.M., Collins, J.G., Stellner, W., McWilliams, R.R., Moreno-Aspitia, A., et al.: Alternative outpatient chemotherapy scheduling method to improve patient service quality and nurse satisfaction. Journal of oncology practice **14**(2), e82–e91 (2018)

[15] Jiang, R., Ryu, M., Xu, G.: Data-driven distributionally robust appointment scheduling over wasserstein balls. arXiv preprint arXiv:1907.03219 (2019)

[16] Kaandorp, G.C., Koole, G.: Optimal outpatient appointment scheduling. Health Care Management Science **10**(3), 217–229 (2007)

[17] Kall, P., Wallace, S.W.: Stochastic programming. Springer (1994)

[18] Kong, Q., Lee, C.Y., Teo, C.P., Zheng, Z.: Scheduling arrivals to a stochastic service delivery system using copositive cones. Operations research **61**(3), 711–726 (2013)

[19] Mak, H.Y., Rong, Y., Zhang, J.: Sequencing appointments for service systems using inventory approximations. Manufacturing & Service Operations Management **16**(2), 251–262 (2014)

[20] Mak, H.Y., Rong, Y., Zhang, J.: Appointment scheduling with limited distributional information. Management Science **61**(2), 316–334 (2015)

[21] Muthuraman, K., Lawley, M.: A stochastic overbooking model for outpatient clinical scheduling with no-shows. Iie Transactions **40**(9), 820–837 (2008)

[22] Shen, S., Wang, J.: Stochastic modeling and approaches for managing energy footprints in cloud computing service. Service Science **6**(1), 15–33 (2014)

[23] Turkcan, A., Zeng, B., Lawley, M.: Chemotherapy operations planning and scheduling. IIE Transactions on Healthcare Systems Engineering **2**(1), 31–49 (2012)

[24] Zacharias, C., Pinedo, M.: Managing customer arrivals in service systems with multiple identical servers. Manufacturing & Service Operations Management **19**(4), 639–656 (2017)