CHEMOTHERAPY SCHEDULING AND NURSE ASSIGNMENT

A Dissertation Presented

By

Bohui Liang

to

The Department of Mechanical and Industrial Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the field of

Industrial Engineering

Northeastern University
Boston, Massachusetts

May 2015
ABSTRACT

Chemotherapy is one of the major treatment methods for cancer patients. Due to increasing need for cancer care, patients experience long waiting times to receive care. A scheduling system that considers resource availabilities and workload balance can reduce patient waiting times and improve patient and staff satisfaction. This research aims to develop advanced chemotherapy scheduling and nurse assignment methods that consider nursing care needs of patients, resource availabilities, and uncertainties in the oncology clinics. The objectives are improving patient flow by reducing patient waiting times, and improving staff workflow by balancing workload and reducing overtime.

Chemotherapy patients usually receive care from multiple resources including oncologists and nurses. The coordination of these services and resources is critical for timely and efficient treatment of patients. We first develop an appointment scheduling method to determine the oncologist and chemotherapy appointments with the objective of balancing the workload for nurses and oncologists. We use discrete event simulation to model the patient flow in the oncology clinic, and show that the proposed scheduling methods can improve patient flow by reducing waiting times and improve staff workflow by providing a more balanced workload.

In oncology clinics, staffing cost is the highest cost after drug costs. It is very important to determine optimal number of nurses in clinics using different care delivery models. We propose multi-objective optimization models to solve nurse assignment problem in functional nursing care delivery model and patient scheduling problem in primary care delivery model. Patient acuity is used to evaluate nurse workload when assigning a nurse to the patient.
Uncertain treatment durations occur due to individual differences between patients such as degree of illness, sensitivity to the treatment, and physical condition of the patient on the treatment day. We develop a two-stage algorithm to solve patient scheduling problem with uncertain treatment durations and multiple resources. We determine the appointment schedules using a mixed integer programming model while considering mean and variance of chair requirements throughout the day. We use a sequential algorithm to determine chair assignments, calculate expected waiting and idle times over all possible realizations of uncertain treatment durations.
ACKNOWLEDGMENTS

First of all, I would like to show my appreciation to my advisor Professor Ayten Turkcan, for her years of support, encouragement and inspiration. I would like to thank her for all her time and effort she has invested on me to help me develop my research and writing skills during my graduate study. I would like to thank her for her patience and kindness to me.

I would like to thank Dr. Ceyhan and Professor Gupta for serving on my committee, listening my presentation, providing me with valuable suggestions and edits to this dissertation. I would like to give special thanks to Dr. Ceyhan for working with me on multiple projects. It is a great pleasure to work with him during my graduate study.

I would like to thank Department of Hematology and Oncology in Lahey Hospital and Medical Center, especially Ms. Jo A. Underhill, Ms. Brenda E. Hill and Dr. Keith Stuart, for their cooperation and valuable advice on the study of improvement of appointment scheduling and patient flow.

I would like to thank all the professors who had taught me and helped me at Northeastern University. Without them, I will not be able to finish my graduate study. I would like to thank all my friends and my colleagues for their support throughout my graduate study. Their companion made my graduate study enjoyable.

Finally, I would like to thank my beloved husband, Zhengxin, for his endless love, support, understanding and caring. I would like to thank my parents, for letting me pursue my educational dreams. I would like to thank them for their love and believing.
# TABLE OF CONTENTS

List of Figures ................................................................. viii

List of Tables ................................................................. xii

1 Introduction ................................................................. 1
   1.1 Research objectives ................................................ 2
   1.2 Summary of research contributions ................................. 4
   1.3 Organization of the dissertation ................................... 6

2 Background and literature review ...................................... 8
   2.1 Chemotherapy treatment in oncology clinics ..................... 9
   2.2 Appointment planning and scheduling .............................. 14
     2.2.1 Appointment planning (dynamic scheduling) over multiple days 14
     2.2.2 Appointment scheduling on a single day ...................... 16
   2.3 Studies that consider patient flow ................................ 19
     2.3.1 Patient flow for chemotherapy treatment ................... 20
     2.3.2 Patient flow for oncologist visit and chemotherapy treatment 21
   2.4 Nurse staffing, scheduling and assignment ....................... 22
     2.4.1 Acuity-based nurse staffing .................................. 23
     2.4.2 Nurse scheduling and nurse assignment ...................... 25
   2.5 Scheduling with uncertainty ......................................... 27
     2.5.1 Sequencing and scheduling with uncertainty in single server settings 28
     2.5.2 Assignment, sequencing and scheduling with uncertainty in multiple server settings 30
   2.6 Summary .............................................................. 33

3 Appointment scheduling and patient flow ............................ 37
5.3.1 Stage 1: Mixed integer programming model to determine appointment time ........................................... 112
5.3.2 Stage 2: An algorithm to determine chair assignment and compute the expected idle time and waiting time ....................................................... 113
5.3.3 Numerical example ......................................................................................................................... 116
5.4 Computational study ....................................................................................................................... 118
5.4.1 Impact of patient mix on appointment schedules ................................................................. 120
5.4.2 Impact of patient mix and number of chairs on performance measures ................................................. 126
5.5 Conclusion ........................................................................................................................................ 136

6 Conclusion ........................................................................................................................................ 138
6.1 Summary ................................................................................................................................. 139
6.2 Future research directions ........................................................................................................... 141

References ........................................................................................................................................... 143

Appendix A Tables of literature review ......................................................................................... 151
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Sample chemotherapy treatment protocol [44]</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Typical patient flow in an oncology clinic</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Patient flow</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Current distribution of oncologist appointment times for each patient type</td>
<td>45</td>
</tr>
<tr>
<td>3.3</td>
<td>Current distribution of chemotherapy appointment times for each patient type</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Proposed distribution of oncologist appointment times for each patient type</td>
<td>48</td>
</tr>
<tr>
<td>3.5</td>
<td>Proposed distribution of chemotherapy appointment times for each patient type</td>
<td>48</td>
</tr>
<tr>
<td>3.6</td>
<td>Number of chairs occupied during the day</td>
<td>50</td>
</tr>
<tr>
<td>3.7</td>
<td>Two-way interaction effect of scheduling method and patient volume</td>
<td>58</td>
</tr>
<tr>
<td>3.8</td>
<td>Two-way interaction effect of (a) scheduling method and arrival time on waiting time to see the provider and (b) arrival time and lab test rate on waiting time to get treatment for type C patients</td>
<td>59</td>
</tr>
<tr>
<td>3.9</td>
<td>Two-way interaction effect of (a) number of resources and patient volume on total working time and (b) number of resources and lab test rate on total working time</td>
<td>60</td>
</tr>
<tr>
<td>3.10</td>
<td>Two-way interaction effect of scheduling method and number of resources on (a) patient waiting time and (b) total clinic working time</td>
<td>61</td>
</tr>
<tr>
<td>3.11</td>
<td>Average total waiting time by appointment time</td>
<td>61</td>
</tr>
<tr>
<td>4.1</td>
<td>Sample schedule</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>(a) Pareto optimal set; (b) Weighted sum method; (c) $\epsilon$-constraint method (adapted from [69])</td>
<td>76</td>
</tr>
</tbody>
</table>
4.3 Algorithm based on $\epsilon$-constraint approach to find nondominated solutions ........................................ 76
4.4 Gantt chart of (a) Nondominated solution 1 (Number of nurses = 3, total waiting time = 14, total overtime = 3), (b) Nondominated solution 4 (Number of nurses = 4, total waiting time = 4, total overtime = 0) .................................................. 80
4.5 Gantt chart of (a) Nondominated solution 1 (Maximum total excess workload allowance $E_s$ is 0 (and 6), total excess workload is 0, total overtime is 2), (b) Nondominated solution 3 (Maximum total excess workload allowance $E_s$ is 6, total excess workload is 7, total overtime is 0) .................................................. 83
4.6 Screenshot of the spreadsheet-based optimization tool for nurse assignment model .................................... 85
4.7 Screenshot of the spreadsheet-based optimization tool for patient scheduling model ................................. 86
4.8 Total overtime for all nondominated solutions for 30 problems in 5, 6 and 7 nurse settings in functional care delivery model ................................. 89
4.9 Total waiting time for all nondominated solutions for 30 problems in 5, 6 and 7 nurse settings in functional care delivery model ................................. 90
4.10 Pareto optimal solutions for problem number 17 with workload 397 .................................................. 90
4.11 Total overtime for all nondominated solutions for 30 problems in (a) 5, (b) 6 and (c) 7 nurse settings in primary care delivery model; solutions are grouped as: solutions found by when 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed, and only when 12 excess workload allowed ........................................ 94
4.12 Total excess workload for all nondominated solutions for 30 problems in (a) 5, (b) 6 and (c) 7 nurse settings in primary care delivery model; solutions are grouped as: solutions found by when 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed, and only when 12 excess workload allowed ........................................ 95
4.13 Pareto optimal solutions for problem number 17 with workload 397; solutions are grouped as: solutions found with 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed ................................. 96
5.1 Scenario generation and probability calculation ................................................................. 114
5.2 Chair assignment, waiting time and idle time calculation for a sample scenario ................................................................. 118

5.3 Appointment schedules for single patient type settings with uniform distribution .......................................................... 122

5.4 Appointment schedules for single patient type settings with truncated normal distribution ................................................. 123

5.5 Appointment schedules for two patient type settings (LL+LH) with (a) uniform and (b) truncated normal distribution ............... 125

5.6 Appointment schedules for two patient type settings (LL+HL) with (a) uniform and (b) truncated normal distribution ............... 126

5.7 Appointment schedules for two patient type settings (LL+HH) with (a) uniform and (b) truncated normal distribution ............... 127

5.8 Appointment schedules for two patient type settings (LH+HH) with (a) uniform and (b) truncated normal distribution ............... 128

5.9 Appointment schedules for two patient type settings (HL+HH) with (a) uniform and (b) truncated normal distribution ............... 129

5.10 Appointment schedules for two patient type settings (LH+HL) with (a) uniform and (b) truncated normal distribution ............... 129

5.11 Appointment schedules for four patient type settings with (a) uniform and (b) truncated normal distribution ......................... 130

5.12 Number of chairs required for four patient type problem with uniform distribution in (a) schedule 1 and (b) schedule 8 ................. 130

5.13 Number of chairs required for four patient type problem with truncated normal distribution in (a) schedule 1 and (b) schedule 5 ...... 131

5.14 Maximum, minimum and average values of (a) expected waiting time and (b) expected idle time for all patient mix and chair settings with uniform service time distribution ........................................... 132

5.15 Maximum, minimum and average values of (a) probability of completion on time and (b) maximum completion time for all patient mix and chair settings with uniform service time distribution ............. 133

5.16 Maximum, minimum and average values of (a) expected waiting time and (b) expected idle time for all patient mix and chair settings with truncated normal service time distribution .................. 134
5.17 Maximum, minimum and average values of (a) probability of completion on time and (b) maximum completion time for all patient mix and chair settings with truncated normal service time distribution . . 135
LIST OF TABLES

2.1 Studies on planning (dynamic scheduling) problem over multiple days 15
2.2 Studies on scheduling problem on a single day .......................... 17
2.3 Studies on patient flow analysis .............................................. 20
2.4 Studies on acuity-based nurse staffing .................................... 23
2.5 Studies on nurse scheduling and nurse assignment ...................... 26
2.6 Studies that solve sequencing problem with uncertain processing times
   in single-server production settings ........................................ 29
2.7 Studies that solve sequencing and scheduling problem with uncertain
   service times in single-server health care settings ....................... 30
2.8 Studies that solve assignment problem with uncertain service times in
   multiple server settings ..................................................... 31
2.9 Studies that solve assignment, sequencing and scheduling problem
   with uncertain service times in multiple server settings ............... 32

3.1 Notation for appointment scheduling model .............................. 46
3.2 Chemotherapy appointment time vs. duration ............................ 50
3.3 Distribution functions used in the simulation model ..................... 52
3.4 Comparison of actual data with simulation output ....................... 54
3.5 Experimental factors ....................................................... 55

4.1 Notation for patient scheduling and nurse assignment problem ........ 68
4.2 Numerical example data for nurse assignment model ................... 77
4.3 Nondominated solutions for nurse assignment in functional care delivery model ......................................................... 78
4.4 Numerical example data for primary nurse model ....................... 81
4.5 Nondominated solutions for patient scheduling for primary care delivery model ................................................................. 82
4.6 Summary of characteristics of patient mix for 30 problems .......... 88
4.7 Nurse skill levels and maximum acuity levels for each nurse .......... 88
4.8 Functional care delivery model: CPU time, number of nondominated solutions, number of infeasible problems, total waiting time and total overtime ................................................................. 91
4.9 Primary care delivery model: CPU time, number of nondominated solutions, number of infeasible problems, total excess workload and total overtime ......................................................... 97
4.10 Advantages and disadvantages of functional and primary care delivery models ................................................................. 101

5.1 Notation for patient scheduling problem with uncertainty ............ 106
5.2 Numerical example data for patients ............................................ 116
5.3 Mixed integer programming model solution for numerical example with $\alpha_1=0.5$, $\alpha_2=0.5$ ................................................................. 117
5.4 Patient types (LL: low mean, low variance, LH: low mean, high variance, HL: high mean, low variance, HH: high mean, high variance) .... 119
5.5 Expected total treatment durations, number of unique schedules, range of MaxER and MaxVarR for single patient type problems ........ 121
5.6 Expected total treatment durations, number of unique schedules, range of MaxER and MaxVarR for two patient type problems ........ 124
5.7 Expected total treatment durations, number of unique schedules, range of MaxER and MaxVarR for four patient type problems ........ 125

A.1 Appointment planning and scheduling studies ............................... 152
A.2 Appointment planning and scheduling studies (cont’d) .................. 153
A.3 Appointment planning and scheduling studies (cont’d) .................. 154
A.4 Appointment planning and scheduling studies (cont’d) .................. 155
A.5 Studies that consider patient flow .............................................. 156
A.6 Studies that consider patient flow (cont’d) .................................. 157
A.7 Studies that consider patient flow (cont’d) .................................. 158
A.8 Studies that develop patient acuity systems for nurse staffing ......... 159
A.9 Nurse scheduling and assignment studies .................................. 160
A.10 Nurse scheduling and assignment studies (cont’d) ....................... 161
A.11 Nurse scheduling and assignment studies (cont’d) ....................... 162
A.12 Nurse scheduling and assignment studies (cont’d) ....................... 163
A.13 Scheduling with uncertainty studies ........................................ 164
A.14 Scheduling with uncertainty studies (cont’d) ............................. 165
Chapter 1

Introduction

According to American Cancer Society [6], half of all men and one third of all women in the U.S. will develop cancer during their lifetimes. Due to increasing need for cancer care, patients have been suffering from long waiting times to receive the care, and the long waiting has become the top reason of patient dissatisfaction [3, 24, 34, 56]. Studies [64, 65] show that the long waiting time is caused by the lack of consideration of resource availability and workload balancing while scheduling patients. A scheduling system that considers resource availability and workload balancing can reduce patient waiting times and improve patient and staff satisfaction.

Chemotherapy is one of the major treatment methods to treat cancer besides surgery and radiotherapy. It is a systemic treatment method that uses drugs to kill cancer cells. Patients usually receive the treatment by intravenous therapy (IV) and treatment duration may range from 30 minutes to 8 hours based on the treatment protocol and patient’s physical condition. Chemotherapy scheduling is to allocate
appointment times to patients considering resource availabilities and treatment durations. Registered nurses (RNs) are responsible for providing the treatment in infusion clinics. Nurse assignment is to assign nurses to patients so that patient waiting times can be reduced, and chemotherapy can be administered safely.

In current practice, appointment scheduling and nurse assignment is made in an ad-hoc manner based on schedulers’ experience and patients’ preferences. Schedulers assume nurses are available if there is an available infusion chair while scheduling appointments. However, if there are multiple patients arriving at the same time, some patients will have to wait due to nurse availabilities, even if there are available infusion chairs. Our research aims to develop advanced chemotherapy scheduling and nurse assignment methods that consider nursing care needs of patients, resource availabilities, and uncertainties in clinics.

In the remainder of this chapter, we will first discuss our research objectives in Section 1.1. The research contributions will be listed in Section 1.2. The outline of the dissertation is provided in Section 1.3.

### 1.1 Research objectives

The first objective of the research is to develop an appointment scheduling method for both oncologist and chemotherapy appointments. Most of the existing studies focus only on chemotherapy appointment scheduling without considering oncologist appointments. However, patients might have both oncologist and chemotherapy appointments on the same day. For those patients, we need to coordinate the appointments so that patients do not wait too much between appointments and delays in one
appointment do not cause idle time for the resource of the next appointment. We propose an integer programming model to determine the oncologist and chemotherapy appointments. The objective is to balance the workload for nurses and oncologists.

The second objective of this research is to test the impact of appointment schedules determined using deterministic optimization model in a clinic environment with several uncertainties including unpunctual arrivals, uncertain treatment durations, add-ons and cancellations. We use discrete event simulation to model the patient flow in an oncology clinic. We test the impact of scheduling methods on clinic performance measures, and show that the proposed scheduling methods can improve patient flow by reducing waiting times and improve staff workflow by providing a more balanced workload.

The third objective of the research is to solve nurse assignment and patient scheduling problems for functional and primary care delivery models, respectively. In current practice, different care delivery models are used in infusion clinics [31]. In functional care delivery model, nurses can be assigned to any patient on the treatment day. In primary care delivery model, the patients can be assigned to their primary nurse only. In clinics that use functional care delivery model, nurse assignment problem is solved every day by the charge nurse. In clinics that use primary care delivery models, patient scheduling problem is solved to find a balanced workload, which dose not exceed primary nurse capacity. Most of the existing studies that solve nurse assignment problem focus on inpatient settings [43, 46, 47, 49, 51, 57, 62]. Our study considers the nurse assignment problem in an outpatient setting where care needs to be provided in a timely manner. In order to balance nurse workload, we use patient acuity system to determine the workload for each patient. Patient acuity is determined based on the care needs and treatment durations. Our study considers
patient scheduling problem for primary care model with the objective of minimizing excess workload and hence reducing the number of part-time nurses. Since staffing cost is the highest cost after drug costs in oncology clinics, the overall aim of solving these two problems is to determine optimal number of nurses in clinics using different care delivery models.

The fourth objective of this research is to solve appointment scheduling problem with uncertain treatment durations considering multiple resource availabilities. We first propose a two-stage stochastic programming model to determine appointment schedules with the objectives of minimizing expected waiting time, idle time and overtime. However, due to the difficulty of solving the proposed stochastic programming model, we propose a mixed integer programming model to determine an appointment schedule while considering mean and variance of chair requirements throughout the day. In the second stage, we determine chair assignments, and calculate expected waiting and idle times while considering all possible realizations of uncertain treatment durations. We test the performance of the proposed algorithm on appointment schedules and corresponding performance measures for different patient mix settings, where patients have different service time distributions.

1.2 Summary of research contributions

The contributions of this study can be summarized as follows:
• We propose a new integer programming model to schedule oncologist and chemotherapy appointments together. The model considers three different types of patients (oncology, chemotherapy, oncology and chemotherapy patients) to balance the workload in the clinic throughout the day. While scheduling patients, the number of nurses in the clinic is taken into consideration to make sure nurses are available when patients arrive for their appointment. The optimal solution is used as a template for patient scheduling in the clinic every day.

• We develop a simulation model of operations in an oncology clinic, featuring uncertainties such as unpunctual arrivals according to appointment times, uncertain treatment durations, multiple patient classes, add-ons and cancellations. It is the first simulation model that includes both consultation and infusion processes with unpunctual patient arrivals. Most of the existing studies use arrival rates instead of appointment schedules in simulation models.

• We propose a multi-criteria optimization model to solve nurse assignment problem in functional care delivery setting. We also propose a multi-criteria optimization model to solve patient scheduling problem in primary care delivery setting. It is the first study that solves acuity-based nurse assignment and patient scheduling problems in an outpatient setting. The proposed models can be used to determine the optimal staffing levels.

• Based on the nurse assignment and patient assignment models, we create spreadsheet tools that solve both problems in Microsoft Excel. The tools can be easily used in the clinics with little training.
• We solve chemotherapy scheduling problem with uncertain treatment durations. To the best of our knowledge, this is the first study that solve scheduling problem considering uncertainty with multiple resource types, multiple servers and unknown sequence. The proposed two-stage algorithm can be used to determine appointment schedules, chair assignments, and calculate the expected waiting time and idle time.

1.3 Organization of the dissertation

The remainder of this dissertation is organized as follows.

In Chapter 2, we first provide a brief introduction of chemotherapy treatment, appointment planning, scheduling and nurse assignment problems in oncology clinics. Then, we present detailed literature review of relevant studies. The review includes studies that solve appointment planning and scheduling problems, model chemotherapy patient flow, solve nurse staffing, scheduling and assignment problems, and solve scheduling problems with uncertainty.

In Chapter 3, the details of our study in an oncology clinic are provided. A mathematical programming model is developed to generate balanced appointment schedules for oncologist visit and chemotherapy treatment. The objective is to balance physician and nurse workload throughout the day. A discrete event simulation model is developed to evaluate the operational performance in the clinic and to identify initiatives for improvement in process flow, scheduling and staffing. It also used to test the impact of the results solved by the mathematical programming model. Our results show
that patient waiting times and clinic total working times can be reduced and more balanced resource utilization can be achieved by using better scheduling methods.

In Chapter 4, we solve nurse assignment and patient scheduling problems in outpatient oncology clinics in order to determine the optimal number of nurses needed. We use integer programming to solve the nurse assignment problem in a functional care setting. By solving the problem, patients’ actual treatment start times and nurse assignment are determined. The objectives are to minimize total patient waiting time between scheduled appointment time and actual start time and to minimize total nurse overtime. We use integer programming to solve patient scheduling problem in a primary care delivery setting. We assume each patient has been assigned to a primary nurse before the appointment is requested. The objectives are to minimize total overtime and total excess workload. We present spreadsheet-based optimization tools in this chapter. The tools are easy to implement in clinics that solve nurse assignment and patient scheduling problems on a daily basis.

In Chapter 5, we solve the appointment scheduling problem with uncertain treatment durations. The objective is to allocate appointment start times for each patient considering their uncertain treatment durations. A mixed-integer programming model is proposed to solve appointment scheduling with the objectives of minimizing maximum expected chair requirement and its variance. After the appointment schedules are determined, an algorithm is used to assign patients to chairs while considering nurse and chair availabilities. The algorithm also calculates expected waiting time and idle time over all realizations of uncertain treatment durations.

In the last chapter, some concluding remarks about the study and future research directions are provided.
Chapter 2

Background and literature review

In this chapter, we first provide a brief introduction about chemotherapy treatment, and the problems solved (chemotherapy planning and scheduling, nurse staffing, scheduling, assignment and scheduling with uncertain service times) in oncology clinics in Section 2.1. Then we provide an extensive review of the literature relevant to our study. Section 2.2 provides a review of the studies that model and solve appointment planning and scheduling problems in oncology and infusion clinics. Section 2.3 includes the studies that consider patient flow to determine resource levels, patient schedules, nurse schedules and clinic layout plans. Section 2.4 is a review of studies that solve nurse staffing, scheduling and assignment problems. Section 2.5 is a review of studies that solve scheduling problem with uncertain service times or job processing times. Section 2.6 provides a summary of our contributions with respect to the existing literature.
2.1 Chemotherapy treatment in oncology clinics

Chemotherapy, which is one of the major methods of treating cancer patients, is a drug therapy that aims to kill cancer cells or stop them from multiplying. However, chemotherapy drugs destroy healthy cells while killing cancer cells. Therefore, chemotherapy treatment is given in cycles, which allows the cancer cells to be attacked at their most vulnerable times, and the normal cells to have time to recover from the damage. Chemotherapy patients visit oncology clinics multiple times during the whole treatment, and the visit frequency depends on treatment protocols. Figure 2.1 is a sample chemotherapy treatment protocol for breast cancer. It shows a 21-day treatment cycle, where patient receives treatment on days 1 and 8 with different medications. The treatment cycle is repeated until disease progression or unacceptable toxicity.

![Chemotherapy Order Template](image_url)

**Figure 2.1:** Sample chemotherapy treatment protocol [44]
In early years, chemotherapy was provided in inpatient settings. Patients needed a hospital bed to receive the treatment. Over the past two decades, chemotherapy treatment shifted from inpatient to outpatient settings due to sophisticated treatment methods and improved management of side effects [66]. As mentioned above, chemotherapy treatment is given in cycles where treatments are followed by a period of rest. Patients only need to come to the clinic for treatment on the treatment days. In order to achieve best results of chemotherapy treatment, patients should receive their treatment on the days noted on treatment protocol. Any delay in treatments will compromise the efficacy [65]. The start date of the treatment and subsequent treatment dates need to be determined before the treatment starts. Chemotherapy appointment planning is determining the dates of patients’ treatment over a planning horizon with the consideration of chair availabilities, nurse availabilities, and treatment durations, with the objective of smoothing workload or balancing number of patients on each day. On the other hand, chemotherapy scheduling is determining the start time of treatment on the treatment day as when patient should arrive to the clinic to receive treatment. Chemotherapy scheduling also includes the allocation of chairs and nurses, to make sure the resources are available at the appointment time. The objective of chemotherapy scheduling is to minimize the clinic overtime, makespan or balance nurse workload. In infusion clinics, each patient’s treatment duration varies from 30 minutes to 8 hours based the drugs and their physical condition, which makes chemotherapy appointment scheduling problem different from other appointment scheduling problems. In most appointment scheduling studies, the appointment duration are usually less variable than chemotherapy treatment [10].

Once chemotherapy patients arrive to the clinic, they need to go through multiple processes before oncology consultation or chemotherapy treatment. A typical patient
flow in an oncology clinic is provided in Figure 2.2. The patient arrives at the clinic based on the appointment time. At the registration or front desk, the patient is checked in. Then, an available Medical Assistant (MA) brings the patient to the vital room to take patient’s vital signs. Vital signs include blood pressure, heart rate, body temperature and weight. Based on oncologist orders, the patient might need a blood test. The blood sample can be taken from vein or PORT. PORT is a small medical appliance implanted under the skin for easy access to the blood stream. It can be used to draw blood, infuse chemotherapy drugs, transfer red blood cells and platelets [21]. Generally, the MA can draw blood sample from patient in a lab room. If a patient has a PORT implanted in the chest, a registered nurse (RN) is required to take blood sample from the PORT. Based on patient’s lab results, oncologist may adjust chemotherapy drugs or dosage. After that, patient proceeds to see the oncologist or receive chemotherapy treatment. Oncology consultation is provided by a designated oncologist in an exam room, and chemotherapy treatment is provided by an RN in an infusion chair. The RN is assigned to the patient before or after patient arrival depending on clinic operations.

As patient flow in Figure 2.2 shows, RNs are the key resources in chemotherapy treatment. In infusion clinics, their responsibility includes patient assessment, patient education, chemotherapy administration (achieving access, premedication, hydration, patient monitoring), management of side-effects, charting, triaging patient questions and problems, and providing counseling to patients and family members [45]. There is high variability in nursing time and nurse workflow due to the treatment protocols that require different infusion methods and treatment durations. The high variability in nursing care needs is caused by patient specific factors such as difficult vein access and risk for hyper-sensitivity reactions. Therefore, patient acuity systems are
Figure 2.2: Typical patient flow in an oncology clinic

developed to determine the nursing care needs. Nurse staffing is determining number of nurses needed in the clinic with the consideration of daily workload in the clinic. Nurse scheduling is determining the specific shift start time and end time for each nurse to make sure there is sufficient number of nurses in the clinic to meet patient needs. Nurse assignment is assigning patients to nurses with the consideration of nurse availabilities, workload, treatment durations, and appointment times.

In oncology clinics, different nursing care delivery models are used for nurse assignment and patient scheduling. In functional care delivery model, nurses are assigned to a group of patients depending on patient mix in a given day. The patients may see different nurses every time they come to the clinic. In primary care delivery model, patients are assigned to a primary nurse and care is provided by the same primary nurse at each visit. In medical care delivery model, nurses assist the physicians as needed and carry out nursing aspects of medical care [31]. According to the Oncology Nursing Society survey, functional and primary care delivery models are the
most commonly used methods (40% use functional care delivery model and 39% use primary care model) in oncology clinics [31]. In this study, we focus on functional and primary care delivery models. In functional care delivery model, acuity-based nurse assignment is to assign nurses to patients considering patients’ acuity levels, nurses’ skill levels and maximum number of patients a nurse can take care of simultaneously in order to balance nurse workload, patient waiting time and clinic overtime. In primary care delivery model, treatment start times need to be determined according to primary nurses’ availabilities to minimize clinic overtime.

In current practice, a patient’s appointment duration depends on chemotherapy infusion time, which is determined based on the treatment regimen, teaching time for new patients, premedications/hydration time [26]. In real clinic environment, each patient’s treatment duration varies due to individual physical condition, degree of illness or sensitivity to the treatment. In this case, scheduling patients based on deterministic treatment durations may cause patient waiting time and nurse idle time in reality. To solve patient scheduling problem with uncertain treatment durations, we need to determine an appointment time for the patients with the allocation of an available chair and a nurse while minimizing expected patient waiting time and nurse idle time for all possible realizations of treatment durations.

There are studies that solve chemotherapy planning and scheduling problem in infusion clinics, studies about patient flow in oncology clinics, studies about patient acuity, nurse staffing, scheduling and assignment and studies that solve scheduling problem with uncertain service time or processing time. The following sections provide a detailed literature review of these studies.
2.2 Appointment planning and scheduling

In Section 2.1.1, we discuss the studies that model and solve appointment planning problems over multiple days. In Section 2.1.2, we discuss the studies that solve appointment scheduling problems on a single day. Tables A.1-A.4 in Appendix provide detailed information including problem solved, objectives, modeling method, and solution approach in appointment planning and scheduling studies.

2.2.1 Appointment planning (dynamic scheduling) over multiple days

The studies that solve the appointment planning problem use optimization methods and heuristics to determine the treatment days according to chemotherapy treatment plans [4, 23, 51, 52, 54, 65]. Table 2.1 provides information on processes (consultation and/or infusion), methods (optimization and/or heuristic), and objectives considered in these studies.

Turkcan et al. [65] is the first study that solves the appointment planning problem for chemotherapy treatment. They propose a mixed integer programming (MIP) model to assign new patients’ treatments to days without changing the plans of existing patients. The objectives are minimization of treatment delays, overutilization and underutilization of resources on each day. They propose a rolling horizon approach to schedule new patients for the treatment. They first solve the model every δ days for a planning horizon of T days (δ < T) for new patients. Then these patients become existing patients in next δ days and their treatment plans remain fixed. Condotta
<table>
<thead>
<tr>
<th>Study</th>
<th>Processes considered</th>
<th>Methods</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkcan et al. (2012) [65]</td>
<td>√ Consultation</td>
<td>√ Infusion</td>
<td>Patient waiting time/Treatment delay</td>
</tr>
<tr>
<td></td>
<td>√ Optimization</td>
<td></td>
<td>Overutilization/Underutilization</td>
</tr>
<tr>
<td></td>
<td>√ Heuristic</td>
<td></td>
<td>Balance daily bed load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clash* density</td>
</tr>
<tr>
<td></td>
<td>√ Optimization</td>
<td></td>
<td>Overutilization/Underutilization</td>
</tr>
<tr>
<td></td>
<td>√ Heuristic</td>
<td></td>
<td>Balance daily bed load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clash* density</td>
</tr>
<tr>
<td>Gocgun and Puterman (2014) [23]</td>
<td>√ Consultation</td>
<td>√ Infusion</td>
<td>Patient waiting time/Treatment delay</td>
</tr>
<tr>
<td></td>
<td>√ Optimization</td>
<td></td>
<td>Overutilization/Underutilization</td>
</tr>
<tr>
<td></td>
<td>√ Heuristic</td>
<td></td>
<td>Balance daily bed load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clash* density</td>
</tr>
<tr>
<td>Sadki et al. (2010a, 2012) [51, 54]</td>
<td>√ Consultation</td>
<td>√ Infusion</td>
<td>Patient waiting time/Treatment delay</td>
</tr>
<tr>
<td></td>
<td>√ Optimization</td>
<td></td>
<td>Overutilization/Underutilization</td>
</tr>
<tr>
<td></td>
<td>√ Heuristic</td>
<td></td>
<td>Balance daily bed load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clash* density</td>
</tr>
<tr>
<td>Sadki et al. (2010b) [52]</td>
<td>√ Consultation</td>
<td>√ Infusion</td>
<td>Patient waiting time/Treatment delay</td>
</tr>
<tr>
<td></td>
<td>√ Optimization</td>
<td></td>
<td>Overutilization/Underutilization</td>
</tr>
<tr>
<td></td>
<td>√ Heuristic</td>
<td></td>
<td>Balance daily bed load</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clash* density</td>
</tr>
</tbody>
</table>

Table 2.1: Studies on planning (dynamic scheduling) problem over multiple days

and Shakhlevich [4] build a multi-level optimization model to generate a scheduling template. The template is used to book new patients’ chemotherapy treatment dates over the planning horizon \(T\) and the appointment times on each day. The objectives of the study are to minimize the waiting time, clash density, and total number of clashes. Clash means a nurse is assigned with multiple tasks in a slot. Gocgun and Puterman [23] use Markov decision process (MDP) to allocate treatment days to patients. Their objective is to minimize the cost of diverting patients and the treatment delays beyond the tolerance limits. Due to the difficulty of solving MDP model, they propose approximate dynamic programming (ADP) and heuristic algorithms.

All above studies [4, 23, 65] consider only the chemotherapy appointment. However, chemotherapy patients need to see their oncologists on certain days of the treatment (i.e. at the beginning of each cycle). Coordinating oncologist appointments with
chemotherapy appointments is important to reduce patient waiting between appoint-
ments. Sadki et al. [51, 54] propose a MIP model to determine both oncologist work
schedules and patient schedules simultaneously with the objective of balancing the
daily bed capacity requirement. In another study [52], they assume that the oncolo-
gist work schedules are given and they use a MIP model to determine treatment days
of new patients without changing the schedule of existing patients.

The studies by Gocgun and Puterman [23] and Sadki et al. [51, 52, 54] propose meth-
ods to determine the days of chemotherapy treatments, but they do not determine the
appointment times on treatment days. Turkcan et al. [65] and Condotta and Shakhle-
vich [4] are the only studies that solve daily appointment scheduling problems after
the treatment days are determined.

2.2.2 Appointment scheduling on a single day

There is significant amount of literature on general appointment scheduling problems
(see Cayirli and Veral [10] for a literature review). However, chemotherapy scheduling
is different from other appointment scheduling studies because chemotherapy treat-
ment durations are more variable than other appointments. Studies [4, 27, 28, 53, 55,
60, 61, 64, 65] discuss appointment scheduling on a single day for chemotherapy treat-
ment and/or oncologist visit. Table 2.2 provides information on processes, methods
and objectives considered in these studies.

The first group of studies that solve scheduling problem on a single day assume the
patients that should be scheduled are known in advance. Turkcan et al. [65] propose
an integer programming model to determine appointment times, nurse and chair
assignments with the objective of minimizing the maximum completion time of all treatments while satisfying nurse and chair availability constraints. Shashaani [61] extends the daily appointment scheduling model of Turkcan et al. [65] by incorporating patient preferences, staggered nurse schedules, and start time constraints (i.e. start after the completion of oncologist appointment). Santibanez et al. [55] propose a multi-objective integer programming model to schedule all patients considering nurse capacity with the objectives of satisfying patients’ time preferences, time constraint from physician schedule, pharmacy capacity, balancing workload between nurses, balancing workload of each nurse throughout the day, and assigning clinical trial patients to specialized nurses.

The second group of studies consider the dynamic arrival of appointment requests and
schedule the patients one at a time as appointments are requested. Hahn-Goldberg et al. [27, 28] use constraint programming to develop a template schedule based on historical data and update the template dynamically when appointment requests do not fit the template. The proposed model determines the start times of drug preparation and treatment with the constraints of pharmacy, nurse, and chair capacities at any time throughout the day. Condotta and Shakhlevich [4] propose an integer programming model to update the daily schedule with the objective of minimizing the work conflicts. Sevinc et al. [60] propose a multiple knapsack model for offline scheduling (where all patients are known). Two heuristics methods for online scheduling are discussed in the study, one is to allocate the patient to the seat which yields highest remaining capacity after the allocation for current patient and the other is to allocate the patient to the seat with minimum remaining capacity. Tanaka [64] uses several bin-packing heuristics that provide a set of rules about how to assign patients to chairs and how to determine an appointment time. Based on these heuristics/rules, the patients are scheduled dynamically upon the appointment requests arrive.

The previous studies [4, 27, 28, 55, 60, 61, 64, 65] consider only the chemotherapy appointments. Sadki et al. [53] determine both oncologist and chemotherapy appointments simultaneously. A mixed-integer optimization model is proposed to determine consultation start times, drug preparation time, and injection start time with the objective of minimizing a weighted combination of patient waiting time and makespan (clinic closing time). The nurses and pharmacists are assumed to have enough capacity so their availability is not considered in the proposed model. A Lagrangian relaxation approach and a local search heuristic are proposed to solve the mixed integer programming model. Similar to [55, 61, 65], Sadki et al. [53] considers a deterministic environment where patient mix is known in advance.
2.3 Studies that consider patient flow

All studies that solve appointment scheduling problem on a single day [27, 28, 53, 55, 60, 61, 64, 65] assume deterministic service times, punctual arrivals, and no uncertainties or delays due to lab, pharmacy or nurse availability as discussed in Section 2.2.2. Most of the appointment scheduling studies do not consider patient flow. However, in real clinic environment, chemotherapy patients go through several processes (check-in, vitals, lab test and infusion treatment) and require multiple resources (check-in staff, MAs, vital rooms, RNs, pharmacists and infusion chairs) at every visit in clinic. Several uncertainties such as variability in service times and patient flow, unpunctual arrivals, add-ons, cancellations, and delays in getting lab results and drug preparation exist in real clinic environments.

The studies that consider patient flow and uncertainties in oncology clinics use discrete event simulation [5, 7, 39, 56, 58, 61, 64, 67, 68] to allocate proper resources, determine operational policies, and find arrival rates for the patients. In Section 2.3.1, the studies that consider patient flow for chemotherapy treatment are reviewed. In Section 2.3.2, the studies that consider patient flow for both oncologist and chemotherapy appointments are reviewed. Table 2.3 provides a summary of processes, methods, and performance measures used in all studies. Tables A.5-A.7 in Appendix provide more detailed information including patient categories, processes, other factors and scenarios considered in each study.
### Table 2.3: Studies on patient flow analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Processes considered</th>
<th>Methods</th>
<th>Objectives / Performance measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed et al. (2011) [5]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Yokouchi et al. (2012) [68]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Woodall et al. (2013) [67]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Shashaani (2011) [61]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Tanaka (2011) [64]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Sepulveda et al. (1999) [58]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Baesler and Sepulveda (2001) [7]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Matta and Patterson (2007) [39]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Santibanez et al. (2009) [56]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

#### 2.3.1 Patient flow for chemotherapy treatment

Simulation modeling in infusion clinics are used to determine nurse schedules and arrival rates in [5, 67, 68]. Ahmed et al. [5] and Yokouchi et al. [68] use simulation to determine the best appointment scheduling rules by changing arrival rates and nurse schedules (number of nurses at each time interval). The objectives are minimization of patient waiting times and maximization of throughput. Woodall et al. [67] use simulation to determine the best daily shift start time with the objective of minimizing patient waiting time in oncology treatment center.

Besides determining the nurse schedules and arrival rates, simulation is also used to determine the impact of different appointment scheduling methods. Shashaani [61] first determines the appointment schedule using a mathematical programming model,
and uses it as an input schedule for the simulation model. She uses simulation to evaluate the impact of service time variability at all stages on key performance measures such as patient waiting time. Tanaka [64] uses simulation to test different appointment scheduling rules generated by bin-packing heuristics. The model is also used to determine the time allocated for pre-treatment processes such as check-in, blood draw and order verification, and to determine the time for preparation and nursing including pharmacy, vitals, assessment, IV or port setup and drug setup process.

2.3.2 Patient flow for oncologist visit and chemotherapy treatment

It is difficult to optimize scheduling and patient flow in oncology clinics without considering the patient flow from upstream stages (i.e. oncologist appointment), because patient flow from upstream stages might incur start time limits, uncertainties (cancellations, add-ons), and delays in downstream stages (i.e. chemotherapy treatment). Therefore, it is important to consider both stages (oncologist and infusion appointments) simultaneously for better coordination of appointment schedules to improve patient flow and balance resource utilization in oncology clinics.

There are three studies [7, 39, 58] that consider both oncologist visit and chemotherapy treatment processes to determine the impact of different resource levels, number of patients scheduled per day, arrival rates, queuing policies, and alternative floor layouts. Sepulveda et al. [58] use discrete event simulation to determine the impact of alternative floor layouts, number of patients scheduled per day, and a new building plan. Baesler and Sepulveda [7] integrate simulation, goal programming and genetic
algorithm to find the best combinations of control variables (i.e., resources) to meet the predetermined goals of patient waiting time, chair utilization, closing time and nurse utilization. Matta and Patterson [39] use simulation to evaluate the impact of different patient arrival rates, resource levels (i.e., additional nurses, doctors), queuing policies, and an express testing center for a group of patients.

There is only one study that considers patient flow for oncology visits. Santibanez et al. [56] use simulation to determine the impact of clinic start time, resident/student involvement, appointment order by patient type, and exam room assignment (dedicated or pooled exam room) on patient waiting time and clinic duration.

### 2.4 Nurse staffing, scheduling and assignment

As discussed in Section 2.1, patient acuity systems are created to evaluate the nursing care requirements. We review the literature that develop patient acuity systems and use the system to determine the optimal nurse staffing levels in Section 2.4.1. In Section 2.4.2, we review literature about nurse scheduling and nurse assignment problems. Nursing scheduling is to determine shift start time and end time for each nurse to minimize nurse hour shortage in the clinic. In order to balance the workload between nurses, nurse assignment models are used to assign nurses to patients with the consideration of workload assigned to each nurse. Table A.8 provides detailed information on studies that develop patient acuity systems and solve nurse staffing problems. Tables A.9-A.12 provide detailed information on nurse scheduling and assignment problems.
2.4.1 Acuity-based nurse staffing

In literature, there are studies that propose patient acuity tools in ambulatory oncology settings to estimate nursing care requirements and determine nurse staffing level [11–13, 20, 29, 32]. Table 2.4 provides a summary of criteria used in development of patient acuity systems and the objectives for developing acuity systems in these studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Acuity evaluation metric</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acuity system</td>
<td>Nurse staffing</td>
<td>Treatment time</td>
</tr>
<tr>
<td>Dobish (2003) [20]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Chabot et al. (2005) [11]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Cusack et al. (2004a,b,c) [12, 13, 32]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Hawley et al. (2009) [29]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Table 2.4: Studies on acuity-based nurse staffing

Dobish [20] categorize chemotherapy protocol into five levels according to the time required to administer the protocol. Similarly, Chabot and Fox [11] develop a patient-classification system that acuity levels are assigned to each regimen based on the number of agents, pre-medications, complexity of administration and assessments required. These two studies do not consider individual differences between patients such
as degree of illness or time needed with patient and/or family members. To determine staffing level, Dobish [20] uses a timetable to schedule next-day chemotherapy patients to each provider according to patient acuity level. In the tabular form, each column represents a provider, the protocol levels and coffee break and lunch break are filled in the cells that reflects nurse schedule. The objective is to schedule as many patients as possible for chemotherapy treatment on the following day of their oncologist appointments. In Chabot and Fox [11], a tabular form is used for nurse-patient assignment. The columns are labeled with the name of each nurse and their shift time, and the rows are patients’ names and acuity levels and nurses’ lunch break.

Considering individual differences, Cusack et al. [12, 13, 32] develop a patient intensity system which is evaluated by nursing time, degree of illness and the complexity of care needed. Nursing leaders use it as a tool to identify the workload and allocate resources properly. On the other hand, Hawley and Carter [29] use total treatment time, time with patient and/or family members, blood draws and any additional nursing needs to determine the acuity level. They provide a scheduling guideline that considers patient acuity levels to determine the number of appointments on each day. Cusack et al. [12, 13, 32] and Hawley et al. [29] use patient acuity system to evaluate number of nurses needed in the clinic. Cusack et al. [12, 13, 32] address that patient intensity level is used to determine nursing time required to see a patient. A number is given as a fraction of 480 minutes (8-hour shift) according to patient intensity. The number of nurses during a day is found by determining the number of patients with different intensity levels. Similarly, Hawley et al. [29] address that in Cleveland Clinic Cancer Center, each nurse is assigned 18 to 24 total patient acuity per day. When the acuity level goes higher than 22, ideally another nurse is needed.
2.4.2 Nurse scheduling and nurse assignment

Nurse scheduling is determining the shift start and end times in outpatient clinics. Mixed integer programming and simulation can be used to determine nurse schedules. Woodall et al. [67] use a mixed integer programming model to find weekly and monthly schedules for different types of nurses. They use simulation to determine the optimal start times of infusion nurses on each day.

Even though there are several studies on personnel staffing and scheduling in operations research literature, there are not many studies that solve nurse assignment problem where patients are assigned to nurses based on their care needs. There are a few studies [43, 46, 47, 50, 57, 62, 63] that focus on acuity-based nurse assignment problem in inpatient setting. In inpatient environment, nurses need to provide care to patients in bed during the shift. Two studies [30, 59] paid their attention to acuity-based nurse assignment problem in home healthcare. In home healthcare, nurses need to travel to patients’ homes to provide care. Table 2.5 provides a summary of methods, setting, and objectives/performance measures considered in these studies.

In inpatient setting, two studies [43, 57] use optimization methods to solve nurse assignment problem. Mullinax and Lawley [43] develop a patient acuity tool and propose an integer linear programming model to assign patients to nurses in a neonatal intensive care unit. The acuity system divides patient care needs into fourteen modules, each module contains multiple levels with different scores. Patient’s acuity is the summation of all care needs from these modules. The neonatal intensive care unit is divided into several zones by location. The nurses can stay only in one zone and cannot be assigned to patients from different zones. The objective of nurse assignment is to balance nurse workload according to patients’ acuity levels. Schaus et
<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Methods</th>
<th>Healthcare setting</th>
<th>Objectives / Performance measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nurse scheduling</td>
<td>Nurse assignment</td>
<td>Integer programming</td>
<td>Constraint programming</td>
</tr>
<tr>
<td>Woodall et al. (2013) [67]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullinax et al. (2002) [43]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schaus et al. (2009) [57]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punnakitikashem et al. (2006, 2008) [46, 47]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenberger (2004) [50]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sundaramoorthi (2009, 2010) [62, 63]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2009) [59]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hertz et al. (2009) [30]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: Studies on nurse scheduling and nurse assignment

al. [57] solve the same problem using constraint programming. They discussed the limitation of using integer programming method, such as it can only solve small size problems, and can only find an approximation of the objective function value since minimizing variance can not be expressed in a linear model.

Three studies [46, 47, 50] aim to minimize excess workload. Punnakitikashem et al. [46, 47] propose a stochastic programming approach that addresses the uncertainty and fluctuations in patient care, and the differences in nursing skills in an inpatient unit. The workload is estimated by direct care and indirect care. Direct care must be performed in a given period and indirect care can be performed throughout the shift. Rosenberger et al. [50] solve nurse-to-patient assignment by integer programming method to minimize nurse excess workload. They classify patients by care needs during day, evening and night.
Sundaramoorthi et al. [62, 63] build a simulation model driven by data mining to evaluate different nurse-to-patient assignment policies. The performance of assignment policies is measured by total assigned care, total unassigned direct care, total direct care, total time spent in non-patient locations and the walking time. The prediction and classification of nurse status is determined by Classification and REgression Trees (CART), and service times are presented by Kernel function.

Kim et al. [59] and Hertz et al. [30] are two studies on nurse-to-patient assignment in home healthcare setting, where nurses visit patients on a regular basis. Both studies build integer programming model to solve nurse-patient assignment problem. Besides workload and time constraints, traveling time is a crucial factor that is considered.

2.5 Scheduling with uncertainty

Two major approaches have been used to solve scheduling problems with uncertainty: stochastic programming and robust optimization. In stochastic programming, the uncertainty is given as distributions. In robust optimization, distribution of the uncertainty is assumed to be unknown, only the upper bound and lower bound of the uncertainty are given. In this section, we will review the studies that propose stochastic programming or robust optimization approaches to solve assignment, sequencing, and scheduling problems with uncertain service times. In section 2.5.1, we discuss the studies solve sequencing and scheduling problem with uncertainty in single server settings. In Section 2.5.2, we review the studies that solve assignment, sequencing and
scheduling problem with uncertainty in multiple server settings. Tables A.13 - 14 provide detailed information on studies that solve scheduling problem with uncertainty using stochastic programming and robust optimization.

### 2.5.1 Sequencing and scheduling with uncertainty in single server settings

The studies that consider uncertain processing times in single machine environment solve sequencing problem using robust optimization. Table 2.6 provides information about methods, system settings and objectives of the studies. Daniels and Kouvelis [15] propose a branch-and-bound algorithm to determine an optimal sequence with the objective of minimizing total flow time. Daniels and Carrillo [14] propose a $\beta$-robust scheduling method to determine the optimal sequence that maximizes the likelihood of achieving flow time performance no greater than a target level. Montemanni [42] proposes a mixed integer programming model to determine job sequence with the objective of minimizing total flow time on a single machine with uncertain processing times. Since all of these studies are performed in production settings, they just determine the optimal sequence. They do not have to determine the scheduled start times of the jobs.

In healthcare environments, due to the importance of patient waiting times, it is not enough to determine the optimal sequence. The start times of the services/procedures (i.e. appointment times) should also be determined a priori. Most of the existing studies that consider uncertain service times solve the sequencing and scheduling problem in surgical settings. Table 2.7 gives detailed information about the studies that solve sequencing and scheduling problems in single-server healthcare settings.
<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Methods</th>
<th>System settings</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniels and Kouvelis (1995) [15]</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Daniels and Carrillo (1997) [14]</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Montemanni (2007) [42]</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

**Table 2.6:** Studies that solve sequencing problem with uncertain processing times in single-server production settings

Denton and Gupta [17] consider a fixed sequence on a single server (operating room), and propose a two-stage stochastic programming model to determine optimal appointment times for a set of surgeries with uncertain durations. The objective of the proposed model is to minimize waiting time, idle time and overtime. Denton et al. [18] propose a stochastic optimization model to solve surgery sequencing and scheduling problem in a single operating room. Since the stochastic mixed-integer-program is NP-hard, they propose several practical heuristics for approximating the optimal solution. Mittal et al. [41] present a robust optimization framework for the appointment scheduling problem in a single server system. In the first part of the study, they allocate service duration for each job with a fixed processing sequence. The objective is to minimize underage cost and overage cost. In the second part of the study, they present several heuristic algorithms to solve the sequencing problem. In our study, we cannot use these approaches, because chemotherapy patients need multiple resources (a chair and a nurse).
2.5.2 Assignment, sequencing and scheduling with uncertainty in multiple server settings

In multiple resource environments, the studies that consider uncertain service times should also solve the assignment problem, where jobs/patients are assigned to resources. Table 2.8 gives the information about the studies that solve the assignment problem with uncertain service time in multiple server environment. Denton et al. [19] propose stochastic programming and robust optimization methods to assign surgeries to operating rooms with the objective of minimizing the fixed cost of opening operating rooms and variable cost of overtime. Rachuba et al. [48] propose a scenario-based robust approach to determine the surgery date and operating room for patients with multiple objectives (minimizing patient waiting times, total amount of overtime and number of patients to be deferred to the next planning period). Min and Yih [40]
generate optimal surgery schedule with uncertain operation durations. The study considers assigning patients to surgical blocks and downstream process (surgical intensive care unit) to minimize surgical block overtime and patient waiting cost. Patient waiting cost is a set of non-decreasing values based on the difference of the scheduled time and the beginning of the planning period. These three studies [19, 40, 48] do not determine the appointment sequence and schedules on the day of the surgery.

Table 2.8: Studies that solve assignment problem with uncertain service times in multiple server settings

<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Methods</th>
<th>System settings</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denton et al. (2010) [19]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rachuba et al. (2013) [48]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Min and Yih (2010) [40]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Keller and Bayraksan [35] and Batun et al. [8] are two studies that solve assignment, sequencing and scheduling problems with uncertain service times in multiple server settings. Table 2.9 gives the detailed information about these two studies. Keller and Bayraksan [35] consider a multi-resource system where the jobs require different amounts of resources at each time period. This problem setting is very similar to our problem where patients require multiple resources at different rates in each time period (a chair is required throughout the treatment and a nurse is required at the beginning of the treatment). Keller and Bayraksan [35] propose a time-indexed mixed integer programming model to find the start times of all jobs while considering the
resource requirements and resource capacities at each time interval. The objective is to minimize the expected total cost related to job start times and penalty cost of exceeding resource capacity. The expected cost is a function of starting time and possible realizations of the processing time (i.e. expected completion time, expected earliness/tardiness). Batun et al. [8] propose a two-stage stochastic programming model to solve assignment, sequencing and scheduling problem in a surgical setting. The first stage decision variables are the number of operating rooms to open, allocation of surgeries to operating rooms, sequence of surgeries in each operating room, and start time of each surgeon. The second stage decision variables are actual completion times, idle time, and overtime for each scenario. The decision variables we consider in our study are allocation of patients to chairs, sequence of patients on each chair, appointment times, actual start times, idle times and overtime, which are similar to the decision variables in Batun et al. [8].

<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Methods</th>
<th>System settings</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Assignment</td>
<td>Sequencing</td>
<td>Scheduling</td>
</tr>
<tr>
<td>Keller and Bayraksan (2009) [35]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Batun et al. (2011) [8]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 2.9: Studies that solve assignment, sequencing and scheduling problem with uncertain service times in multiple server settings
2.6 Summary

Chemotherapy planning and scheduling have attracted sufficient attention from researchers. We have discussed multiple studies on appointment planning and scheduling. We also looked at studies that use simulation to analyze patient flow and determine optimal resource level, nurse schedule or patient schedule. We discussed studies about nurse staffing problem based on patient acuity system, nurse scheduling and assignment problems. At the end of this chapter, we also discussed studies that solve assignment, sequencing and scheduling problems considering uncertain service times in healthcare settings and uncertain processing times in production settings.

In chemotherapy planning and scheduling, we focus on chemotherapy scheduling problem with the coordination of oncology appointment scheduling. Instead of solving one day problem, we use patient mix from historical data as the input and find the proportion of each appointment duration in each time slot. The proposed schedule can be used as a scheduling template to determine patient appointments dynamically. Different from Sadki et al. [53], which the objectives are to minimize patient waiting time and clinic closing time, our objective is to distribute appointments evenly throughout the day to balance the workload in the clinic for both oncologists and nurses.

Considering patient flow in oncology clinics, we develop a discrete event simulation model with multiple patient types, various patient routing, resource requirements, and unpunctual patient arrivals, uncertain service times. The simulation model is used to test the impact of our chemotherapy scheduling method considering both oncologist visit and chemotherapy treatment processes besides different operational decisions (patient arrival times, lab rates and resource levels). Instead of testing different
patient arrival rates like [39], we test the impact of appointment schedule determined by an optimization model. In our model, patient appointments are generated from a template which is created from the solution of the optimization model. In our study, we consider multiple patient classes with varying routings and resource requirements, unpunctual arrivals, and stochastic service times and treatment durations.

To determine nurse staffing level in infusion clinics, we solve nurse assignment problem and patient scheduling problem under functional and primary care delivery model using patient acuity to evaluate nurse workload. Different from studies [11–13, 20, 29, 32], which use patient acuity systems to balance nurse workload and determine optimal nurse staffing level, we solve acuity-based nurse assignment and patient scheduling problem to determine nurse staffing level by evaluating patient waiting time, nurse overtime and excess workload required. All of the existing studies solve the nurse assignment problem in inpatient setting [43, 46, 47, 50, 57, 62, 63] or in home healthcare setting [30, 59]. To the best of our knowledge, this is the first study that solves nurse assignment problem for a given patient mix and appointment schedule in an outpatient setting. The difference between inpatient setting and outpatient setting is that, in outpatient setting, nursing care needs to be provided in a timely manner. In outpatient clinics, workload balancing and treatment start time are considered when assigning nurses to patients. In real practice, most nurse assignments are either based on judgment of the charge nurse or same number of patients are assigned to each nurse to provide similar caseload [65]. Even though there are appointment scheduling studies that assign patients to nurses while determining the appointment times [61, 65], our study is different from these studies. We focus on both functional and primary care delivery model in outpatient setting. In functional care delivery model, we solve nurse assignment problem considering patient acuity,
nurse skill and treatment duration. We also determine treatment start time according to nurses’ availability while patients’ appointment times are pre-determined. In primary care delivery model, since each patient will be assigned to their primary nurses, we determine treatment start time according to nurses’ availability in order to minimize nurse overtime.

We consider assignment, sequencing, and scheduling problem in a multiple resource environment to solve patient scheduling problem with uncertain treatment durations. The studies by Keller and Bayraksan [35] and Batun et al. [8] are the most relevant studies to our study. In Keller and Bayraksan [35], the objective is to minimize the expected total cost related to job start times and penalty cost of exceeding resource capacity. The expected cost is a function of starting time and possible realizations of the processing time (i.e. expected completion time, expected earliness/tardiness), which can be calculated without knowing the sequence of jobs. However, in our study, we cannot calculate the waiting time of patients without knowing the sequence of patients in all chairs. Therefore, we cannot use a time-indexed integer programming model similar to Keller and Bayraksan [35] without addition of new decision variables to determine the optimal sequence on each chair. In Batun et al. [8], they first determine number of operating rooms to open, allocation of surgeries to operating rooms, sequence of surgeries in each operating room, and start time of each surgeon then determine actual completion times, idle time, and overtime for each scenario. In our study, we do not have to make the chair assignments (i.e. operating room assignments in [8]) and determine the sequence of patients on each chair (i.e. sequence of surgeries in each operating room in [8]) in the first stage. This is due to the easiness of changing the chair assignments to reduce patient waiting times when actual treatment durations are realized. Therefore, in a stochastic programming model, our
first stage decision variables would be the appointment times of the patients. The chair assignments, actual start times, idle times, and overtime would be the second stage decision variables, which are determined when the actual treatment durations are realized. Our study is the first study that solves scheduling problem with the consideration of uncertain treatment durations to minimize expected waiting time, idle time and overtime. It is the first study that considers assignment, sequencing and scheduling problem over multiple servers and multiple types of resources in infusion clinics.
Chapter 3

Appointment scheduling and patient flow

Cancer patients often receive multiple treatments including chemotherapy, radiotherapy, and surgery from different specialists for extended periods of time. Services such as blood work, physical exam, drug preparation, and chemotherapy administration, are required to be performed in different facilities such as laboratories, clinics, pharmacies, and treatment rooms. The services in each facility are performed by multiple resources such as phlebotomists, nurses, pharmacists, and medical oncologists. Clinic administrators face the difficult decision of improving efficiency in this complex multi-facility environment. The coordination of these services and resources is critical for timely and efficient treatment of patients. In this study, our aim is to show that delays due to inefficient care delivery can be eliminated by better coordination, planning, and scheduling.
3.1 Introduction

This study is performed in the Department of Hematology and Oncology in Lahey Hospital and Medical Center, Burlington MA. We consider chemotherapy patients who come to the clinic according to their appointment times for oncologist visit and/or chemotherapy treatment, and go through multiple stages (oncologist visit, lab, pharmacy, chemotherapy administration). They require multiple resources (oncologists, nurses, chairs, pharmacists, phlebotomists) at each stage of the process. Uncertainties such as unpunctual arrivals, delays in laboratory and pharmacy areas, increase or decrease in treatment durations due to side effects or dose changes, cancellations, and add-ons, occur during a typical clinic day. All these uncertainties affect patient flow and staff workflow. Patients experience long waiting times due to delays in laboratory, pharmacy, and chemotherapy administration areas, and providers and staff experience an unbalanced workload throughout the day.

Reducing patient waiting times is among the highest priorities for quality improvement and patient satisfaction in outpatient cancer treatment facilities [22]. Appointment scheduling that does not consider the availability of clinic resources and nursing care requirements is determined to be the main cause of delays and unbalanced workload [11, 25]. In this study, our aim is the incorporate the actual resource requirements into coordination and scheduling of appointments to minimize patient waiting times and balance clinic workload.

We use discrete event simulation to model the patient flow in the oncology clinic and test the impact of different operational decisions on patient waiting times, resource utilizations and overtime. The model considers multiple patient classes with varying routings and resource requirements, unpunctual arrivals, and stochastic service times
and treatment durations. Earlier simulation studies, which proposed changing the arrival rates to have a smoother workload, did not develop a method to find a schedule that considers the dependencies between oncologist and chemotherapy appointments. In this study, we propose an optimization model to determine a coordinated appointment schedule for oncology and infusion clinics with the objective of balancing workload and resource utilization for both processes during the day. This study is one of the few studies that considers all the complexities and uncertainties that occur in multi-facility healthcare systems with multiple patient classes and varying patient routings.

The contributions of this study are:

1. In this study, we use optimization and simulation approaches to improve chemotherapy patient flow and scheduling in an outpatient oncology clinic. We develop a mathematical programming model that evenly distributes patients into time slots to balance the workload throughout the day for oncologist and chemotherapy appointments. Instead of determining a scheduling template, we use the optimal schedule to determine a probability matrix that shows the probability of assigning different patients types (categorized according to their treatment durations) to different appointment times. With the probability matrix, the scheduler does not have to know the whole day demand and can schedule the patients as they arrive sequentially.

2. We develop a discrete event simulation model that closely mimics the complex flow of chemotherapy patients in a real clinic environment. The simulation model incorporates several environmental complexities including unpunctual
arrivals, stochastic oncologist and chemotherapy appointment durations (functions of scheduled appointment durations), add-ons, cancellations, and nurse workflow. It also considers multiple patient classes characterized by appointment types, laboratory test requirement, treatment durations, oncologist visit durations, and nursing times. The simulation model is used to evaluate the current performance and test alternative operational decisions to improve performance measures such as patient waiting times and clinic overtime.

The remainder of this chapter is organized as follows. The clinic environment including patient flow, patient mix, and clinic resources is explained in detail in Section 3.2. In Section 3.3, the appointment scheduling in current practice and proposed optimization model to find a balanced appointment schedule are explained. The details of the simulation model and the computational results are given in Sections 3.4 and 3.5. The last section provides concluding remarks.

3.2 Clinic Environment

We worked with The Hematology and Oncology Clinic at Lahey Hospital and Medical Center in Burlington, MA. The clinic provides care for patients with blood disorders and cancer. Patients come to the clinic for consultation or follow-up with the oncologist and for chemotherapy treatment. Patients arrive to the clinic according to their appointment times and go through several processes before being seen by the medical oncologist and/or receive chemotherapy treatment.
3.2.1 Patient flow

Figure 3.1 shows the patient flow in the hematology and oncology clinic. When patient arrives to the clinic, appointment scheduling coordinator (ASC) verifies patient information, prints medication list, and updates patient status in oncology information system. After check-in, a medical assistant (MA) prepares patient’s chart. When the vital room becomes available, MA calls the patient from waiting room and takes vital signs. If the patient needs laboratory tests, then blood is drawn by the medical assistant in the lab room. However, if the patient has a PORT, the blood can be drawn only by a registered nurse (RN) in the infusion clinic. The patient waits in the waiting room until an RN becomes available for blood draw. The blood sample is sent to the central lab and the patient waits until the lab results are received. The patients who need lab tests are told to arrive one hour early to the clinic to have enough time for blood draw and laboratory tests. After the test results are received, MA takes the patient who has an oncologist appointment to an available exam room and notifies the oncologist. If the patient has both oncologist and infusion appointments, he/she receives chemotherapy after the oncologist appointment. If the patient has a chemotherapy appointment, RN assesses patient condition (test results, vital signs, and drug dose) before the treatment can start. If the patient’s health status is good enough to receive the treatment on that day, he/she is seated on a chemotherapy chair and pharmacy is informed for drug preparation. When the drug is ready, RN picks up the drug and starts the treatment. While the patient is on the chair, RN continuously monitors patient status. When the treatment is completed, RN finishes the treatment and patient is discharged. The dotted lines in Figure 3.1 show the time stamps we get from the information system. We use these time stamps to determine patient mix, scheduled and actual consultation and infusion durations, appointment
schedules, arrival times, patient waiting times, and time in system for each patient type.

3.2.2 Patient mix

The patients are first divided into three groups based on the appointments they have on a given day: i) patients with oncologist appointment only (Type O); ii) patients with chemotherapy appointment only (Type C); and iii) patients with both oncologist and chemotherapy appointments (Type OC). The oncologist appointment durations change according to provider practice, and whether patient is a new or an existing patient. The oncologists allocate 10 to 30 minutes for existing patients and 40 to 60 minutes for new patients. The chemotherapy treatment durations depend on chemotherapy protocols and show a high variability (range from 30 to 360 minutes). Based on current patient mix, approximately 45% of infusion appointments are scheduled for 30 minutes. The patients who are scheduled for their first chemotherapy treatment require additional nursing time for education. In order to allocate more nursing time to new patients, we determined whether the patient is a new patient or not. Patients who need lab tests require additional resources. Therefore, we further classified each patient group as i) patients who need lab tests; and ii) patients who need lab tests and have PORT access.
PT: patient; ASC: appointment scheduling coordinator; MA: medical assistant; RN: registered nurse; OC: doctor and infusion appointment; O: doctor appointment only; C: infusion appointment only.

**Figure 3.1: Patient flow**
3.3 Appointment scheduling

3.3.1 Current practice

Patient access to the oncology clinic is guaranteed through appointments. The patients who have to see their oncologists and receive chemotherapy on the same day are scheduled based on oncologist availability and chemotherapy treatment duration. Long chemotherapy treatments are scheduled at earlier times to avoid overtime. When we look at the existing schedules, we observe that more patients of type OC are scheduled for oncologist appointments in the morning (Figure 3.2) to provide enough time for chemotherapy afterwards. Majority of oncologist appointments are scheduled between 8:30 AM and 4:00 PM. Figure 3.3 shows the percentage of patients scheduled for chemotherapy treatment. Chemotherapy appointments are scattered throughout the day with no single peak time. However, more appointments are scheduled in the middle of the day compared to the rest of the day. The current schedules create unbalanced workload in the clinic and we believe a better scheduling method can provide a more balanced workload and lower patient waiting times.

3.3.2 Proposed scheduling method

We propose a mathematical programming model to find a better schedule with a balanced workload. Table 3.1 shows the notation used in the proposed model.

The following model assumes patient mix is given, where the number of appointments for each patient type and appointment duration are known. The first objective (1.a)
minimizes the difference between maximum and minimum number of chairs occupied to find a balanced chair utilization. The second objective (1.b) minimizes the difference between maximum and minimum number of exam rooms occupied at each time slot. Constraints (2) and (3) make sure all patients are scheduled for chemotherapy and oncologist appointments. When the patient has both oncologist and chemotherapy appointment, there should be enough slack time between the appointments, which
is satisfied by constraint (4). Constraints (5) and (6) determine the number of chairs and exam rooms occupied at each time slot. Chemotherapy treatment can start only when a nurse is available and constraint (7) is used to limit the number of treatment starts based on the number of nurses. Constraints (8) and (9) are the capacity constraints for chairs and providers. Constraint (10.a) and (10.b) are the integrality constraints.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$NC_{id}$</td>
<td>Total number of patients of type $i$ with chemotherapy appointment duration $d$ ($i = OC, C$)</td>
</tr>
<tr>
<td>$NO_{jd}$</td>
<td>Total number of patients of type $j$ with oncologist appointment duration $d$ ($j = OC, O$)</td>
</tr>
<tr>
<td>$R_t$</td>
<td>Number of nurses available at time slot $t$</td>
</tr>
<tr>
<td>$P_t$</td>
<td>Number of physicians available at time $t$</td>
</tr>
<tr>
<td>$T$</td>
<td>Total number of slots in planning horizon</td>
</tr>
<tr>
<td>$F$</td>
<td>Number of chairs</td>
</tr>
<tr>
<td>$s$</td>
<td>Slack time required between oncologist and chemotherapy appointments</td>
</tr>
<tr>
<td>$x_{idt}$</td>
<td>Number of patients of type $i$ with appointment duration $d$ scheduled to start chemotherapy at time $t$</td>
</tr>
<tr>
<td>$y_{jdt}$</td>
<td>Number of patients of type $j$ with appointment duration $d$ scheduled to start oncologist visit at time $t$</td>
</tr>
<tr>
<td>$c_t$</td>
<td>Number of chairs occupied at time slot $t$</td>
</tr>
<tr>
<td>$o_t$</td>
<td>Number of oncologist appointments scheduled to start at time $t$</td>
</tr>
</tbody>
</table>

**Table 3.1: Notation for appointment scheduling model**

\[
\begin{align*}
\min & \quad \max_t c_t - \min_t c_t \\
\min & \quad \max_t o_t - \min_t o_t \\
\text{s.t.} & \quad \sum_{t=1}^{T-d+1} x_{idt} = NC_{id} \quad i = OC, C \text{ and } \forall d \quad (2) \\
& \quad \sum_{t=1}^{T-d+1} y_{jdt} = NO_{jd} \quad j = OC, O \text{ and } \forall d \quad (3)
\end{align*}
\]
\[ \sum_{d} y_{OC,d,t} = \sum_{d} x_{OC,d,t+s}, \quad t = 1 \cdots T - s \] (4)

\[ \sum_{d} \sum_{u=\max\{t-d+1,1\}}^{t} (x_{OC,d,u} + x_{C,d,u}) = c_t, \quad t = 1 \cdots T \] (5)

\[ \sum_{d} \sum_{u=\max\{t-d+1,1\}}^{t} (y_{OC,d,u} + y_{O,d,u}) = o_t, \quad t = 1 \cdots T \] (6)

\[ \frac{\sum_{d}(x_{OC,d,t} + x_{C,d,t})}{\sum_{d}(NC_{OC,d} + NC_{C,d})} \leq \frac{R_t}{\sum_{t=1}^{T} R_t}, \quad t = 1 \cdots T \] (7)

\[ c_t \leq F, \quad t = 1 \cdots T \] (8)

\[ o_t \leq P_t, \quad t = 1 \cdots T \] (9)

\[ x_{idt}, c_t \geq 0 \text{ and integer, } \quad i = OC, C \text{ and } \forall d, t = 1 \cdots T \] (10.a)

\[ y_{jdt}, o_t \geq 0 \text{ and integer, } \quad j = OC, O \text{ and } \forall d, t = 1 \cdots T \] (10.b)

We solved the proposed model in two stages. In the first stage, we solved the model with the first objective (1.a) and constraints (2), (5), (7), (8) and (10.a) to determine the number of chemotherapy appointments that should be scheduled at each time slot. Once chemotherapy appointments are determined, the number of oncologist appointments for the patients who have both appointments are given as inputs to the second model with the second objective (1.b) and constraints (3), (4), (6), (9) and (10.b) to determine the number of oncologist appointments for other patients. Figures 3.4 and 3.5 shows the proposed schedules for oncologist and chemotherapy appointments, respectively. According to the optimal oncologist appointment schedule, the last two appointment slots are allocated to type O patients since those patients do not need chemotherapy treatment on the same day. For chemotherapy appointments, most
type C patients are scheduled in early morning because their appointments do not need to be coordinated with the physician schedule.

Figure 3.4: Proposed distribution of oncologist appointment times for each patient type

Figure 3.5: Proposed distribution of chemotherapy appointment times for each patient type

Even though the proposed mathematical programming model is supposed to be solved as an integer programming model, we solve the problem as a linear programming model and use the results to determine a probability matrix. The probability matrix
is used to cope with the sequential nature of scheduling process, where patients are scheduled one at a time and the appointment time is determined based on the appointment duration and predetermined probability of scheduling the appointment at a given time slot. Table 3.2 shows the optimized matrix used for proposed scheduling practice. Once there is an appointment request, a random number between 0 and 1 will be assigned to it. The chemotherapy appointment time is determined based on the random number and appointment duration. For example, for a 270 minutes appointment, if the random number is smaller than 0.5241, the appointment time will be 7:30am. If it is between 0.5241 and (0.5241+0.0313), the appointment time will be 8:00am and so on. If the patient also has an oncology appointment, we have to make sure there is available oncology appointment slot before the chemotherapy appointment. Otherwise, a new random number will be generated to find a new chemotherapy appointment and oncology time until we find the available time slot for the appointments. Figure 3.6 shows the chair utilization for the current practice and proposed scheduling method. The proposed scheduling method gives a smoother workload during the day compared to current practice. The difference between the optimal and generated schedule is due to the randomness in using a probability matrix and sequential appointment schedule generation.

3.4 Simulation Model

We developed a simulation model of the clinic to identify the problems related to patient flow in current practice and evaluate the impact of a balanced appointment schedule on key operational measures including patient waiting times, clinic total working time, and resource utilizations.
### Table 3.2: Chemotherapy appointment time vs. duration

<table>
<thead>
<tr>
<th>Time</th>
<th>30</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>150</th>
<th>180</th>
<th>210</th>
<th>240</th>
<th>270</th>
<th>300</th>
<th>330</th>
<th>360</th>
<th>390</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 AM</td>
<td>0.0006</td>
<td>0.0023</td>
<td>0.0000</td>
<td>0.1216</td>
<td>0.2046</td>
<td>0.0079</td>
<td>0.3137</td>
<td>0.5241</td>
<td>0.1686</td>
<td>0.0276</td>
<td>0.0266</td>
<td>0.0498</td>
<td></td>
</tr>
<tr>
<td>8:00 AM</td>
<td>0.0987</td>
<td>0.0024</td>
<td>0.0017</td>
<td>0.0082</td>
<td>0.0111</td>
<td>0.0079</td>
<td>0.2049</td>
<td>0.0013</td>
<td>0.2088</td>
<td>0.0406</td>
<td>0.2650</td>
<td>0.9502</td>
<td></td>
</tr>
<tr>
<td>8:30 AM</td>
<td>0.0843</td>
<td>0.0024</td>
<td>0.0017</td>
<td>0.0082</td>
<td>0.0104</td>
<td>0.0032</td>
<td>0.1216</td>
<td>0.0015</td>
<td>0.0616</td>
<td>0.1766</td>
<td>0.3137</td>
<td>0.2488</td>
<td></td>
</tr>
<tr>
<td>9:00 AM</td>
<td>0.0868</td>
<td>0.0024</td>
<td>0.0003</td>
<td>0.0076</td>
<td>0.0042</td>
<td>0.0538</td>
<td>0.0291</td>
<td>0.0015</td>
<td>0.1185</td>
<td>0.2744</td>
<td>0.3492</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:30 AM</td>
<td>0.0907</td>
<td>0.0024</td>
<td>0.0003</td>
<td>0.0002</td>
<td>0.0033</td>
<td>0.0466</td>
<td>0.0295</td>
<td>0.0680</td>
<td>0.1353</td>
<td>0.2039</td>
<td>0.2305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>0.0930</td>
<td>0.0024</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0038</td>
<td>0.0551</td>
<td>0.2046</td>
<td>0.0682</td>
<td>0.1038</td>
<td>0.2490</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30 AM</td>
<td>0.1039</td>
<td>0.0022</td>
<td>0.0000</td>
<td>0.0158</td>
<td>0.0117</td>
<td>0.0015</td>
<td>0.0989</td>
<td>0.0238</td>
<td>0.0748</td>
<td>0.1070</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>0.0476</td>
<td>0.0095</td>
<td>0.0001</td>
<td>0.0083</td>
<td>0.0088</td>
<td>0.0166</td>
<td>0.1024</td>
<td>0.0204</td>
<td>0.2243</td>
<td>0.0015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:30 AM</td>
<td>0.0029</td>
<td>0.0211</td>
<td>0.0000</td>
<td>0.0422</td>
<td>0.0966</td>
<td>0.0568</td>
<td>0.0617</td>
<td>0.0031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>0.0611</td>
<td>0.0864</td>
<td>0.0000</td>
<td>0.0805</td>
<td>0.1505</td>
<td>0.1928</td>
<td>0.1754</td>
<td>0.0007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:30 PM</td>
<td>0.0491</td>
<td>0.0742</td>
<td>0.0000</td>
<td>0.0891</td>
<td>0.1413</td>
<td>0.1613</td>
<td>0.4423</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00 PM</td>
<td>0.0744</td>
<td>0.0226</td>
<td>0.0770</td>
<td>0.0579</td>
<td>0.1885</td>
<td>0.0214</td>
<td>0.0038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:30 PM</td>
<td>0.0749</td>
<td>0.1159</td>
<td>0.0733</td>
<td>0.0000</td>
<td>0.3137</td>
<td>0.0154</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00 PM</td>
<td>0.0051</td>
<td>0.0216</td>
<td>0.0000</td>
<td>0.6279</td>
<td>0.0082</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:30 PM</td>
<td>0.0026</td>
<td>0.1803</td>
<td>0.0517</td>
<td>0.0088</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:00 PM</td>
<td>0.0155</td>
<td>0.4157</td>
<td>0.0079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:30 PM</td>
<td>0.0429</td>
<td>0.0026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Number of chairs occupied (proportion)](image)

**Figure 3.6:** Number of chairs occupied during the day

### 3.4.1 Input data

We consider the patient mix in current practice, where 34% of patients have both oncologist and chemotherapy appointment (Type OC patient), 45% of patients have oncologist appointment only (Type O patient), and 21% of patients have chemotherapy appointment only (Type C patient). The percentages of patients who need laboratory tests are 56%, 9% and 19% for Type OC, Type O and Type C patients,
respectively. The percentages of patients who need laboratory test and have PORT access are 71%, 11%, and 83%, respectively.

We consider unpunctual arrivals and uncertain service times. Table 3.3 shows all the distributions used in the simulation model. In order to determine the arrival times, the difference between the appointment time and arrival time is calculated. However, since the patients who need lab tests are asked to arrive one hour early to their appointment to have enough time for lab tests, we fitted different arrival time distributions for each patient type. For example, the patients who have oncologist appointment only (type O patients) and need lab test arrive on the average 72.9 minutes early to their oncologist appointment.

The oncologist appointment and chemotherapy treatment durations might be longer or shorter than the scheduled durations due to several reasons. For example, difficulty in intravenous (IV) access or side effects of drugs might increase treatment durations. The side effects of chemotherapy drugs might lead to cancellations after the treatment starts. To consider the actual durations in the simulation model, we fitted distributions for the ratio between actual and scheduled durations. We use different distributions for short infusions (treatment duration ≤ 60 minutes) and long infusions (treatment duration > 60 minutes). For example, if a patient is scheduled for a 30-minute chemotherapy treatment, the actual treatment duration will be $30 \times 2.33 \times Beta(1.21,2.7)$ minutes. That means, the actual treatment duration changes between zero and 70 minutes and the expected value is 21.6 minutes ($21.6 = 30 \times 2.33 \times 0.72$ where 0.72 is the expected value for $Beta(1.21,2.7)$) for 30 minute appointments.
### Table 3.3: Distribution functions used in the simulation model

<table>
<thead>
<tr>
<th>Processes</th>
<th>Distribution</th>
<th>Fitted/Expert Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arrival time - appointment time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type OC patient with lab</td>
<td>Normal(63.44)</td>
<td>Fitted (p=0.0484)</td>
</tr>
<tr>
<td>Type OC patient without lab</td>
<td>Normal(50.54)</td>
<td>Fitted (p=0.133)</td>
</tr>
<tr>
<td>Type O patient with lab</td>
<td>Normal(73.42)</td>
<td>Fitted (p&gt;0.15)</td>
</tr>
<tr>
<td>Type O patient without lab</td>
<td>Normal(29.41)</td>
<td>Fitted (p=0.015)</td>
</tr>
<tr>
<td>Type C patient with lab</td>
<td>Normal(60.31)</td>
<td>Estimated</td>
</tr>
<tr>
<td>Type C patient without lab</td>
<td>Normal(19.60)</td>
<td>Fitted (p=0.0139)</td>
</tr>
<tr>
<td><strong>Actual / scheduled duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncologist appointment</td>
<td>Lognormal(0.068,0.502)</td>
<td>p&gt;0.15</td>
</tr>
<tr>
<td>Chemotherapy (scheduled ≤ 60 min)</td>
<td>2.33 * Beta(1.21, 2.7)</td>
<td>p&gt;0.15</td>
</tr>
<tr>
<td>Chemotherapy (scheduled &gt; 60 min)</td>
<td>1.57 * Beta(1.25, 1.6)</td>
<td>p&gt;0.15</td>
</tr>
<tr>
<td><strong>Other service times</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check-in</td>
<td>Triangular (0.5, 1, 2)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>Time to get the chart ready</td>
<td>Lognormal (1.019, 0.716)-2.5</td>
<td>Fitted (adjusted) (p=0.267)</td>
</tr>
<tr>
<td>Taking vital signs</td>
<td>Triangular(3, 5, 10)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>Blood draw in lab</td>
<td>Erlang(3.98,2)+0.5</td>
<td>Fitted (p=0.024)</td>
</tr>
<tr>
<td>Blood draw in infusion room</td>
<td>Triangular(2.5, 11.6, 36.5)</td>
<td>Fitted (p&gt;0.75)</td>
</tr>
<tr>
<td>Lab turnover time</td>
<td>Triangular(5, 15, 30)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>RN assesses patient condition</td>
<td>Triangular(1,2,10)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>Pharmacy time</td>
<td>Weibull(10.5, 1.42)-1.5</td>
<td>Fitted (adjusted) (p&gt;0.75)</td>
</tr>
<tr>
<td>RN starts chemo (new patient)</td>
<td>Triangular(25,30,45)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>RN starts chemo (established patient)</td>
<td>Triangular(5,10,15)</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>RN finishes chemo</td>
<td>Triangular(2,5,10)</td>
<td>Expert opinion</td>
</tr>
</tbody>
</table>
We included all other stages of the patient flow process including registration/check-in, taking vitals, blood draw, lab turnover time, pharmacy time for drug preparation, and nursing time to start and finish chemotherapy. In order to determine the service time distributions at these stages, we collected data from the oncology information system, performed additional time studies and received expert opinion. The other parameters (three ASCs, six MAs, twelve oncologists, two pharmacists, nine RNs and eighteen infusion chairs) are assumed to be fixed and used as inputs in the simulation model except number of RNs and chairs will be changed as one of the experiment factors in Section 3.6.

3.4.2 Discrete event simulation model

We used Anylogic simulation software to model the patient flow in the oncology clinic. The patients are generated according to the current patient mix in the clinic as explained in Section 3.2. The appointment times are determined based on the scheduling method. For the patients who have two appointments, chemotherapy appointment time is determined first to reduce overtime. In current practice, the schedulers leave a 30-minute gap between oncologist and chemotherapy appointments (chemotherapy appointment time ≥ oncologist appointment time + oncologist appointment duration + 30 minutes).

For verification of the simulation model, we performed statistical analysis to compare the simulation outputs with the real data. We took 100 replications for five days and compared the results with the real data collected over five days. Table 3.4 shows the confidence intervals for the simulation model and the real system data. The results
show that there is no significant difference between the results, which confirms the validity of the simulation model.

<table>
<thead>
<tr>
<th>Performance measures</th>
<th>Actual Mean</th>
<th>95% CI</th>
<th>Simulated Mean</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wait time to see a physician</td>
<td>11.70</td>
<td>[10.06, 13.35]</td>
<td>11.71</td>
<td>[11.49, 11.93]</td>
</tr>
<tr>
<td>Wait time to have treatment (Type OC)</td>
<td>11.18</td>
<td>[7.59, 14.77]</td>
<td>8.49</td>
<td>[8.20, 8.79]</td>
</tr>
<tr>
<td>Wait time to have treatment (Type C)</td>
<td>19.07</td>
<td>[12.86, 25.28]</td>
<td>20.49</td>
<td>[19.94, 21.04]</td>
</tr>
<tr>
<td>Time in system</td>
<td>132.23</td>
<td>[121.47, 142.76]</td>
<td>123.46</td>
<td>[121.57, 125.36]</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of actual data with simulation output

3.5 Computational study

We consider five experimental factors to show the impact of appointment scheduling on clinic performance in a clinic environment with several complexities including multiple processes and resources, unpunctual arrivals, delays, add-ons and cancellations. The factor levels can be seen in Table 3.5. These five factors can affect patient waiting time, total time in the system and clinic working time directly. Those are the performance measures of patient flow.

The first factor is the patient volume, which changes between 80 and 120 patients per day. Higher patient volume will cause longer waiting time. If patients are poorly scheduled, the effectiveness will be more significant. The second factor is the mean difference between appointment times and arrival times. In current practice, the patients who need laboratory tests are asked to arrive one hour early for their appointments. Based on our analysis, we identified that other patients who do not need laboratory test arrive 30 minutes early for their appointment. We consider three
<table>
<thead>
<tr>
<th>Factors</th>
<th>Level I</th>
<th>Level II</th>
<th>Level III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient volume</td>
<td>80 patients/day</td>
<td>100 patients/day</td>
<td>120 patients/day</td>
</tr>
<tr>
<td>Mean difference between appointment time and arrival time (patients with lab, patients without lab)</td>
<td>(60, 30) minutes</td>
<td>(45, 15) minutes</td>
<td>(30, 5) minutes</td>
</tr>
<tr>
<td>Percentage of patients who need lab test</td>
<td>Current rate</td>
<td>20% increase</td>
<td>40% increase</td>
</tr>
<tr>
<td>Appointment scheduling method (AS)</td>
<td>Current AS</td>
<td>Proposed AS</td>
<td>Proposed AS</td>
</tr>
<tr>
<td>Nurse schedule (NS)</td>
<td>Staggered NS</td>
<td>Staggered NS</td>
<td>Non-staggered NS</td>
</tr>
<tr>
<td>Number of chairs and nurses</td>
<td>15 chairs, 8 nurses</td>
<td>18 chairs, 9 nurses</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5: Experimental factors

factor levels to determine the impact of the mean difference between appointment time and arrival times on patient waiting times. The first factor level of (60, 30) corresponds to current practice. The second factor level of (45, 15) and third factor level of (30, 5) assume patients who need laboratory test arrive on the average 45 and 30 minutes early to the clinic, respectively. The other patients arrive on the average 15 and 5 minutes early to the clinic, respectively.

The third factor is the percentage of patients who need laboratory tests. Although the chemotherapy protocols determine the need for laboratory tests, the patients might choose to have the test at another location and have the results sent to the clinic. The percentage of patients who need laboratory test might increase when patients prefer to have the test on the day of appointment in the clinic. As the percentage of patients with laboratory test increases, the workload and patient waiting times increase. The current percentages of patients who need tests are 55%, 9% and 19% for type OC, type O and type C patients, respectively. We considered 20% and 40%
increase with respect to current percentages as the other factor levels.

The fourth factor is the appointment scheduling method and nurse schedules. In current practice, staggered nurse schedules are used where nurses start working at different times. According to the current appointment scheduling method used in the clinic, less number of patients are scheduled in early morning hours, which shows that staggered nurse schedules are taken into consideration. The proposed appointment scheduling method has the potential to provide a more balanced workload throughout the day. However, when a staggered nurse schedule is used, the number of available nurses should be considered in determining the optimal schedule. In order to match available number of nurses and chemotherapy appointments in each time slot, less patients are scheduled in the early morning hours. In order to see the impact of scheduling more patients in the early morning hours, we also consider non-staggered nurse schedules with the optimal appointment scheduling method. The non-staggered nurse schedule assumes same starting times for all nurses and the proposed appointment scheduling method reduces the clinic total working time by scheduling more patients in the early morning hours. The fifth factor is the number of oncology nurses (RNs) and number of staffed chairs. The clinic currently has 9 nurses and 18 chairs for infusion patients. However, the number of nurses might be reduced due to sick calls. We assume it is not safe to treat more patients when number of nurses is less. So we consider staffed chairs, which should be reduced when there are less number of RNs available. In order to measure the impact of decrease in number of nurses and staffed chairs on patient waiting time and clinic total working time, we use 8 nurses and 15 chairs as the first factor level and 9 nurses and 18 chairs as the second factor level.

Jun et al. [33] present high patient throughput, low waiting times, a short length of
stay at clinic and low clinic overtime indicate an effective and efficient patient flow. In
our study, the clinic is appointment based that patient throughput is not considered.
The performance measures used for comparison are patient waiting time to see the
provider, patient waiting time for chemotherapy treatment, total waiting time, and
total working time. Patient waiting time to see the provider, and for chemotherapy
treatment are calculated with respect to appointment time and arrival time. If a
patient arrives early, the waiting time before the appointment time is not included
in the calculations. However, chemotherapy patients who need lab tests are asked
to arrive early for their appointments. They go through several processes and they
wait between the processes due to limited resources. Instead of just looking at the
waiting times from appointment time, we consider the total waiting time to show the
importance of coordination between stages. Total working time shows the difference
between the time last patient leaves the system and the clinic start time.

3.5.1 Results

The simulation model is run for a single day and replicated 100 times for each factor
combination, resulting in 16200 (3 × 3 × 3 × 3 × 2 × 100) simulation runs. We
performed ANOVA to analyze the main and interaction effects of all factors on the
performance measures. The main effects of five factors are found to be significant
for all performance measures which means: a) the proposed scheduling method can
reduce patient waiting time, patient time in system and total working time; b) patients
do not need to arrive one hour earlier for their lab or 30 minutes earlier for their
appointments. A shorter time difference between arrival and appointment time can
reduce patient waiting time and total time in the system. Figures 3.7 - 3.9 show the selected significant interaction effects on performance measures.

Figures 3.7.a and 3.7.b show the interaction effect of scheduling method and patient volume on total waiting time and total working time, respectively. The proposed scheduling method gives lower patient waiting time compared to the current scheduling method. The effect of the proposed scheduling method is more significant when the patient volume is high. That means, using a more balanced schedule becomes more important especially when the workload is high. When the patient volume is low (80 patients/day), the waiting times are close to each other for all scheduling methods. The proposed scheduling method gives lower total working time compared to the current scheduling method. The proposed scheduling method with staggered nurse schedule gives 21 minutes lower total working time compared to the current method. The non-staggered nurse schedule gives 14 minutes lower total working time compared to the staggered nurse schedule since more patients are scheduled in the early morning hours.

![Interaction plot for total waiting time](image1)

![Interaction plot for clinic total working time](image2)

**Figure 3.7:** Two-way interaction effect of scheduling method and patient volume on (a) patient waiting time and (b) clinic total working time
Figure 3.8.a shows the interaction plot between the scheduling method and the arrival times. The proposed algorithm gives lower patient waiting time to see the provider compared to the current practice. The same figure shows that the waiting time to see the provider is 15 minutes for the patients who arrive (60,30) minutes early to their appointment and it increases to 21 minutes if they arrive (45,15) minutes early. However, we would like to note that the waiting times are calculated from the appointment time, and do not include the time from arrival to appointment time. If we include the waiting time due to early arrival, the patients who arrive much earlier than their appointment time would actually end up waiting more than other patients who arrive later. For example, when the patients arrive (45,15) minutes early, even though the waiting time increases by 6 minutes (from 15 minutes to 21 minutes) compared to patients who arrive (60,30) minutes early, they save 15 minutes of waiting time, which results in 9 minutes lower waiting time. The interaction plot between the arrival time and the laboratory test rate shows that the arrival time is critical especially when more patients need laboratory tests (see Figure 3.8.b).

**Figure 3.8:** Two-way interaction effect of (a) scheduling method and arrival time on waiting time to see the provider and (b) arrival time and lab test rate on waiting time to get treatment for type C patients.
Figures 3.9.a and 3.9.b show the effect of number of resources on total working time. When the number of resources increases, the decrease in clinic total working time is higher for high patient volume and high percentage of patients who need laboratory tests. Figures 3.10.a and 3.10.b show the interaction effect of scheduling method and number of resources on waiting time for chemotherapy treatment and total working time, respectively. When the clinic has less resources, the proposed algorithm gives a higher improvement in patient waiting time for Type C patients compared to current practice (7 minutes (22%) improvement for 15 chairs, 8 nurses and 4 minutes (15%) improvement with 18 chairs, 9 nurses). The effect of using the proposed scheduling algorithm with staggered nurse schedule instead of current appointment schedule on total working hours is higher when number of resources is high. The effect of using a non-staggered schedule instead of a staggered schedule has a higher effect on total working time when number of resources is low.

The main objective of the proposed scheduling method is to provide a more balanced workload throughout the day. Figure 3.11 shows the average waiting time by appointment time. The results show that the proposed algorithm reduces the waiting
Figure 3.10: Two-way interaction effect of scheduling method and number of resources on (a) patient waiting time and (b) total clinic working time.

Figure 3.11: Average total waiting time by appointment time.
3.5.2 Summary

As a summary, the proposed scheduling method gives better clinic performance, especially when patient volume is high. The non-staggered nurse schedule provides a lower clinic total working time due to higher number of patients scheduled in early morning hours. The arrival time of the patients is a critical factor that affects total waiting times. In current practice, patients who need laboratory tests are asked to arrive one hour early, which increases the patient waiting time unnecessarily. We showed that a shorter time between the arrival time and appointment time can reduce the waiting times for patients. The clinics should determine the time allocated for laboratory tests according to lab turnover times and the percentage of patients who need laboratory tests. The number of available resources (number of chairs and nurses) can become a critical factor that affects waiting times and clinic total working times especially when the patient volume is high.

3.6 Conclusion

A discrete event simulation model was developed to model the complex flow of chemotherapy patients in oncology clinics. The model considers multiple patient classes with varying routings and resource requirements, unpunctual arrivals, uncertainties in service and treatment durations, add-ons, and cancellations. The model is used to evaluate the performance of a real clinic and to test alternative operational decisions to improve system performance. Earlier simulation studies, which proposed changing the arrival rates to have a smoother workload, did not develop any scheduling method to find a schedule that considers the dependencies between oncologist and
chemotherapy appointments. In this study, we developed an optimization model to determine a coordinated appointment schedule for oncology and infusion clinics. The proposed scheduling method determines the number of oncologist and chemotherapy appointments with the objective of minimizing the deviation between low and high utilization time slots. The proposed scheduling method is tested using the simulation model and is shown to reduce patient waiting times at peak hours.

The computational results showed that scheduling methods that aim to balance the workload provides lower patient waiting times and clinic total working times. Using a better scheduling method becomes more important especially when patient volume is high. The operational decisions that are not determined based on the actual data can cause unnecessary waiting times. For example, asking the patient to arrive one hour early can cause high waiting times, especially when the lab turnover time is much less.

We presented our results to clinic managers and they decided to implement the scheduling method to reduce patient waiting times during peak hours. The clinic managers decided to do an analysis of lab turnover times to understand the effect of patient volume throughout the day and determine the optimal time that should be allocated for the patients who need laboratory tests to minimize patient waiting times. The clinic also started scheduling the last patient half hour early so that overtime can be reduced. Patient acuity can be considered in future research while scheduling patient. Patient acuity is used to quantify the needs of nursing care and nurse workload. Panel size study can be another future research topic to determine the optimal nurse-patient ratio based on patient acuity and nurse workload.

The operational difficulties in oncology clinics are common to any other healthcare
system where patients are seen in different departments/clinics on the same day and require a large number of resources (physicians, nurses, pharmacists, technicians, medical assistants). Multiple patient classes, varying patient routings, day-of-week and time of day differences in patient volume and resource availabilities complicate the process of improving the efficiency of these multi-facility systems. This study is one of the few studies that considers all the complexities and uncertainties that occur in multi-facility healthcare systems.
Chapter 4

Acuity-based nurse assignment and patient scheduling

Chemotherapy patient scheduling and nurse assignment are complex problems due to high variability in treatment durations and nursing care requirements. Nurses are the key resources that provide chemotherapy treatment in oncology clinics. According to the ONS survey, 41% of the responding nurses were responsible for scheduling patients and 54% were fixing scheduling problems [31]. This shows the complexity of appointment scheduling in infusion clinics, because valuable nursing time is used for patient scheduling. In this study, we focus on functional and primary care delivery models, and propose optimization methods to reduce the time spent for nurse assignment and patient scheduling. We believe patient acuity systems can estimate the nursing requirements for each patient more accurately. The integration of acuity systems, nurse workflow, and patient scheduling can provide better schedules that minimize patient waiting times and staff overtime, and balance workload for the nurses.
4.1 Introduction

As mentioned in Chapter 2, there are two types of nursing care delivery models used in oncology clinics, functional and primary care delivery models. In this study, we focus on these two models, and propose optimization methods to reduce the time spent for nurse assignment and patient scheduling. The aim is to determine the optimal number of nurses with the objectives of minimizing patient waiting times and nurse overtime in functional care delivery model and minimizing excess workload and nurse overtime in primary care delivery model.

The contributions of this study are:

1. This is the first study that solves nurse assignment problem for a given patient mix and appointment schedule in an outpatient setting. Due to predetermined appointment schedules and staff schedules with fixed start and end times, timeliness is important in outpatient settings. The proposed multiobjective optimization model finds schedules that minimize total patient waiting time and clinic overtime simultaneously.

2. This is the first study that considers primary care delivery model in oncology clinics. The clinics, which use primary nurse model to improve continuity of care, might experience high variability in daily nurse workload due to treatment protocols. The proposed model finds several schedules that minimize total overtime and total excess workload simultaneously. The proposed model can be used as a decision making tool to determine the number of part-time nurses required when the workload is higher than the primary nurses’ capacity.
3. The proposed methods can reduce the time spent for daily nurse assignment and patient scheduling tasks significantly. Two spreadsheet-based optimization tools, which use open-source optimization software (OpenSolver), are developed for easy implementation. The developed tools require minimal training and can be used as decision making tools to determine the optimal staffing levels required for safe chemotherapy treatment.

In the remainder of this section, the problem is defined with its underlying assumptions in Section 4.2. In Section 4.3, two multiobjective optimization models are proposed to solve the nurse assignment problem for the functional care delivery model, and the patient scheduling problem for the primary care model. A numerical example and spreadsheet based optimization tools are explained in the same section. Computational results along with managerial insights are discussed in Section 4.4, and concluding remarks are provided in Section 4.5.

### 4.2 Problem definition

In this study, we consider patient scheduling and nurse assignment problems in outpatient oncology clinics using functional and primary care delivery modes, respectively. The notation that will be used throughout the paper can be seen in Table 4.1.

We consider outpatient oncology clinics where fixed start times and regular working hours are commonplace. The clinics set their regular working hours according to patient demand and volume, and availability of providers. The outpatient clinics might run from 7am to 5pm, and provide longer hours on certain days of the week.
Parameters:
\( S \) Number of slots
\( P \) Number of patients
\( D_i \) Treatment duration of patient \( i \)
\( N \) Set of nurses
\( H_{j}^{s}, H_{j}^{f} \) Work schedule (shift start and end times) of nurse \( j \)
\( L_{i} \) Acuity level of patient \( i \)
\( K_{j} \) Skill level of nurse \( j \)
\( n_{ij} \) Takes value 1 if the skill level of nurse \( j \) is enough to treat patient \( i \)
\( M_{j} \) Maximum acuity level for nurse \( j \)
\( A_{i} \) Appointment time of patient \( i \) (for functional delivery model only)

Decision variables:
\( y_{ijs} \) Binary variable, 1 if patient \( i \) is treated by nurse \( j \) and the treatment starts at time slot \( s \)
\( t_{i} \) Treatment start time of patient \( i \) (for functional delivery model only)
\( w_{i} \) Waiting time of patient \( i \) (for functional delivery model only)
\( o_{j} \) Overtime of nurse \( j \)
\( e_{js} \) Excess workload of nurse \( j \) at time slot \( s \)
(for primary care delivery model only)

Table 4.1: Notation for patient scheduling and nurse assignment problem

to accommodate patient demand (i.e. patients who work can come to the clinic after work). The day is divided into smaller time slots (i.e. 30 minutes) and patients are scheduled to arrive at the beginning of these pre-determined slots. The patient may need more than one slot according to the treatment duration and these time slots are blocked once the patient is scheduled.

We consider a single stage system where \( P \) patients are scheduled only for the infusion appointment. The laboratory tests and oncologist appointments that occur before the infusion appointment are not considered. The pharmacy time for chemotherapy preparation is assumed to be included in the treatment duration (\( D_{i} \)). The treatment durations, which might range between 30 minutes and 8 hours, are assumed to be
given. We assume punctual arrivals where patients come to the clinic for chemotherapy treatment at their appointment times.

Nurses are the key resources who administer chemotherapy to patients. Based on clinic working hours, nurses might have different start times and end times. For example, a nurse with 8-hour schedule might start at 7am and work until 3pm, and another nurse with 10-hour schedule might start at 8am and work until 6pm. This type of nurse schedule (staggered nurse schedule) is commonly used in outpatient settings to adjust the availability of nurses according to changing demand throughout the day and provide flexible working hours for the nurses. We assume a staggered nurse schedule with $H^s_j$ and $H^f_j$ as the starting and ending time of working hours for nurse $j$. If the patients are still being treated or waiting for the treatment at the end of the shift, the nurse who provides the service will continue working to complete the treatment.

A nurse is assigned to multiple patients for administering the chemotherapy. The assignment is made based on nurse working hours, skill level of nurse, patient acuity, and maximum number of patients a nurse can simultaneously treat. Each patient has an acuity ($L_i$) level, which represents the complexity of the treatment and the nursing time required. Nurses are assigned to the patients based on their skill level ($K_j$). A nurse can be assigned to a patient only if her skill level is higher than the patient acuity ($n_{ij} = 1$). We assume a nurse can treat multiple patients at the same time. The maximum acuity level ($M_j$) defines how many patients a nurse can treat simultaneously. For instance, nurse $j$ can treat patient $p_1$ and $p_2$ whose acuity levels are 2 and 3 at the same time if her maximum acuity level is greater than or equal to 5. We also assume a nurse can start at most one treatment in each slot.
A sample schedule with three patients and one nurse is provided in Figure 4.1 to clarify the notation. The skill level and maximum acuity level of the nurse are 3 and 5, respectively. The shift start and end times are 3 and 16. Even though the appointment time of the first patient is 1, the treatment cannot start until time 3 due to the shift start time of the nurse. Patient 2 has to wait until next time slot, because a nurse cannot have more than one treatment start in any time slot. Patient 3 is scheduled to arrive at slot 10. However, the treatment cannot start until slot 12 due to maximum acuity level of 5. That means, the nurse cannot take care of two patients with acuity level 3 at the same time.

Figure 4.1: Sample schedule
4.3 Proposed optimization models

4.3.1 Functional care delivery model: Multiobjective optimization model for nurse assignment

We propose a multiobjective optimization model with the objectives of minimizing patient waiting time and nurse overtime. The proposed model assigns nurses to patients and determines the actual treatment start times of the patients. We assume patient schedules are given with appointment times \((A_i)\), treatment durations \((D_i)\) and acuity levels \((L_i)\). The nurse work schedules \((H^x_j, H^f_j)\), skill levels \((K_j)\) and maximum acuity level a nurse can handle at any given slot \((M_j)\) are also given.

Objectives: We consider objectives of minimizing total patient waiting time \((O.1)\) and nurse overtime \((O.2)\).

\[
\begin{align*}
\min \quad TWT &= \sum_{i=1}^{P} w_i = \sum_{i=1}^{P} [t_i - A_i] = \sum_{i=1}^{P} \left[ \sum_{j \in N} \sum_{s=1}^{S} (s - 1)y_{ij}s - A_i \right] \quad (O.1) \\
\min \quad TOT &= \sum_{j \in N} o_j \quad (O.2)
\end{align*}
\]

Assignment constraints: The proposed model aims to allocate a nurse to each patient and determine the start time of the treatment. The decision variable \(y_{ij}s\) takes value 1 when nurse \(j\) is assigned to patient \(i\) and the treatment starts in time
slot $s$. Constraint (1.a) ensures that each patient is assigned to only one nurse who has
enough skill to treat the patient. The treatment can start after the nurse assigned to
the patient starts working for the day and the patient arrives for his/her appointment
$(s \geq \max\{A_i, H_j^s\} + 1)$. Due to the intensity of tasks that should be performed at
the beginning of the treatment, a nurse can start at most one treatment in any given
slot, which is guaranteed by constraint (2.a).

$$\sum_{j \in N} \sum_{s = \max\{A_i, H_j^s\} + 1}^{S} (n_{ij} \times y_{ijs}) = 1 \quad i = 1 \cdots P \quad (1.a)$$

$$\sum_{i = 1}^{P} (n_{ij} \times y_{ijs}) \leq 1 \quad \forall j \in N \quad (2.a)$$

$$s = \max\{A_i, H_j^s\} + 1 \cdots S$$

**Acuity constraints:** Nurses can treat limited number of patients simultaneously
due to nursing requirements and safety issues. Constraint (3.a) makes sure the total
acuity level of patients assigned to a nurse does not exceed the maximum acuity level.

$$\sum_{i = 1}^{P} \sum_{u = \max\{1, s - D_i + 1\}}^{s} (n_{ij} \times L_i \times y_{iju}) \leq M_j \quad \forall j \in N, s = 1 \cdots S \quad (3.a)$$

**Nurse overtime:** Nurse overtime is the difference between the treatment completion
time of the last patient assigned to the nurse and end of regular working hours for
that nurse. Constraint (4.a) calculates the overtime for each nurse.

$$o_j \geq n_{ij} \times y_{ijs} \times (s + D_i - 1) - H_j^f \quad i = 1 \cdots P, \forall j \in N, s = 1 \cdots S \quad (4.a)$$
Non-negativity and integrality: The non-negativity (5.a) and integrality (6.a) constraints make sure all variables are non-negative and $y_{ijs}$ is binary.

$$w_i, o_j \geq 0 \quad i = 1 \cdots P, \forall j \in N$$

(5.a)

$$y_{ijs} \in \{0, 1\} \quad i = 1 \cdots P, \forall j \in N, s = 1 \cdots S$$

(6.a)

4.3.2 Primary care delivery model: Integer programming model for patient scheduling

We propose a multiobjective optimization model to solve the patient scheduling problem. We assume the primary nurse for each patient is known, and appointments are scheduled based on the primary nurse availability. The proposed integer programming model is as follows:

$$\min \ TOT = \sum_{j \in N} o_j$$

(O.2)

$$TEW = \sum_{j \in N} \sum_{s=1}^{S} e_{js}$$

(O.3)

$$\text{st} \quad \sum_{s=H_{t_i}+1}^{S} y_{i,r_i,s} = 1 \quad i = 1 \cdots P$$

(1.b)
\[ \sum_{i=1}^{P} y_{ijs} \leq 1 \quad \forall j \in \mathbf{N}, s = H_j + 1 \ldots S \] (2.b)

\[ \sum_{i=1}^{P} \sum_{u=\max\{1,s-D_i+1\}}^{s} (L_i \times y_{iju}) \leq M_j + e_{js} \quad \forall j \in \mathbf{N}, s = 1 \ldots S \] (3.b)

\[ o_j \geq y_{ijs} \times (s + D_i - 1) - H_i^f \quad i = 1 \ldots P, j \in \mathbf{N}, s = 1 \ldots S \] (4.b)

\[ \sum_{j \in \mathbf{N}} e_{js} \leq E_s \quad s = 1 \ldots S \] (5.b)

\[ o_j \geq 0, e_{js} \geq 0 \quad \forall j \in \mathbf{N}, s = 1 \ldots S \] (6.b)

\[ y_{ijs} \in \{0, 1\} \quad i = 1 \ldots P, \forall j \in \mathbf{N}, s = 1 \ldots S \] (7.b)

The model determines the appointment times for all patients while minimizing the total excess workload (TEW) and total overtime (TOT) simultaneously. Different from the functional care delivery model, the patients can only be assigned to their primary nurse \( r_i \) (constraint 1.b), and the total workload assigned to each nurse at each slot can exceed the maximum acuity level (constraint 3.b). The decision variable \( e_{js} \) is added to the right-hand-side of constraint (3.b) to calculate the excess workload for each nurse at each slot. However, since high workload can cause patient safety problems, we assume part time nurses can be used to take the excess workload. Even though the proposed model does not assign patients to specific part-time nurses, it restricts the total excess workload at each slot (constraint 5.b), where the upper bound \( (E_s) \) is determined according to the maximum acuity level part-time nurses
can handle. Similar to functional care delivery model, a nurse can start at most one treatment at each slot (constraint 2.b), and constraint (4.b) calculates the overtime for each nurse. Constraints (6.b) and (7.b) are non-negativity and integrality constraints for the proposed model.

The proposed nurse assignment and patient scheduling models have multiple objectives and our aim is to find the nondominated solution set that minimizes all objectives simultaneously. For a minimization problem, solution $y$ is said to dominate $y'$ if $f_i(y) \leq f_i(y')$ for all objectives ($i \in \{1, 2, \ldots, n\}$) and $f_i(y) < f_i(y')$ for at least one objective $i$. In Figure 4.2.a, solutions A, B, C, and D form the nondominated solution set which minimizes both $f_1$ and $f_2$. The other solutions (E, F, G, and H) are dominated by the solutions in the nondominated solution set.

Different approaches are used to solve multiobjective optimization problems in the literature. The weighted sum method uses a weighted linear combination of the objectives and assigns different weight combinations to determine the set of nondominated solutions. The $\epsilon$-constraint method converts $k - 1$ of the $k$ objectives into constraints and finds nondominated solutions by changing the right-hand-sides (upper bounds) of these constraints. Figures 4.2.b and 4.2.c show how each method works to find the nondominated solutions.

We use $\epsilon$-constraint method to solve our optimization problems. We convert the overtime objective into a constraint ($\sum_{j \in N} o_j \leq \epsilon$), and then solve the models with different $\epsilon$ values to find the nondominated solutions. Figure 4.3 shows the algorithm used to generate all nondominated solutions for the functional care delivery model, where the objective is minimization of total waiting time. The algorithm for the
primary care model uses the objective of minimization of total excess workload instead of total waiting time.

**Figure 4.2:** (a) Pareto optimal set; (b) Weighted sum method; (c) $\epsilon$-constraint method (adapted from [69])

**Figure 4.3:** Algorithm based on $\epsilon$-constraint approach to find nondominated solutions
4.3.3 Numerical example

In this section, we give a small numerical example to show the schedules generated by the proposed integer programming models. We consider 20 chemotherapy patients to be seen in one day. The day is divided into 16 half-hour slots and treatment durations range from 1 to 9 slots (e.g. 30 minutes to 4.5 hours). The patient acuities range from 1 to 3. Table 4.2 shows the appointment times, durations and acuity levels for each patient. We assume there are at most 4 nurses scheduled for the day. Their skill levels are 3, 3, 2, and 2, and maximum acuity levels are 6, 5, 5, and 4, respectively. We consider an outpatient clinic with regular working hours from 8am (slot 0) to 4pm (slot 16). All nurses start working at time slot 0 and regular working hours end at slot 16.

<table>
<thead>
<tr>
<th>Patient number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment time</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Appointment duration</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Acuity level</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Patient number</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Appointment time</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Appointment duration</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Acuity level</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.2: Numerical example data for nurse assignment model

Nurse assignment model:

First, we solve the proposed multiobjective optimization model with different staffing levels (3 and 4 nurses) to find the optimal nurse assignment and actual treatment start times. Table 4.3 shows the objective function values of the nondominated solutions found by $\epsilon$-constraint approach. When there are three nurses, the nurse assignment model gives two nondominated solutions. The first solution gives total waiting time
of 14 slots and total overtime of 3 slots. The second solution has higher waiting time (16 slots) and lower overtime (1 slots).

<table>
<thead>
<tr>
<th>Nondominated solution</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nurses</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Obj: Total waiting time (slots)</td>
<td>14</td>
<td>16</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Obj: Total overtime (slots)</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Waiting time per patient (slots) (min, average, max)</td>
<td>0, 0.70, 6</td>
<td>0, 0.80, 6</td>
<td>0, 0.15, 1</td>
<td>0, 0.20, 4</td>
</tr>
<tr>
<td>Overtime per nurse (slots) (min, average, max)</td>
<td>0, 1.00, 2</td>
<td>0, 0.30, 1</td>
<td>0, 0.25, 1</td>
<td>0, 0.00, 0</td>
</tr>
</tbody>
</table>

Table 4.3: Nondominated solutions for nurse assignment in functional care delivery model

Table 4.3 also shows the minimum, average, maximum values for waiting time and overtime in the last two columns. These minimum and maximum values are presented to show the range of waiting time and overtime for individual patients and nurses, respectively. For the first nondominated solution, the average waiting time per patient is 0.7 slots, that is 21 minutes (14 slots × 30 minutes/slot / 20 patients = 21 minutes/patient). The minimum and maximum waiting times are 0 and 6 slots (0 and 180 minutes). The average overtime is 1 slot (30 minutes) per nurse (3 slots × 30 minutes/slot / 3 nurses = 30 minutes/nurse). The minimum and maximum overtime are 0 and 2 slots (0 and 60 minutes).

Even though patient waiting time and staff overtime are the most commonly used performance measures in appointment scheduling literature, other performance measures such as resource utilizations are also important in infusion clinics. Since nurse skill levels and maximum acuity levels that can be handled by each nurse differs, we use the following formula to calculate the nurse utilizations.
\[ u_j = \frac{\sum_{i=1}^{P} \sum_{s=1}^{S} (y_{ijs} \times D_i \times L_i)}{(H_{sj} - H_{fj}) \times M_j} \quad \forall j \in N \]

The numerator calculates the total acuity for all patients assigned to a nurse and the denominator calculates the maximum acuity a nurse can handle during regular working hours. For nondominated solution 1, the nurse utilizations are 94\%, 98\% and 81\% for nurses 1–3, respectively. For nondominated solution 4, the nurse utilizations are 86\%, 60\%, 69\%, and 70\%, for nurses 1–4, respectively.

The resource utilizations calculated using the above formula is an average value for the day and it does not show how the workload changes throughout the day. In order to see the workload variation throughout the day, we should look at the number of patients and total acuity assigned to each nurse in each time slot. Figure 4.4 shows the patients assigned to each nurse, treatment start times and durations for two nondominated solutions: nondominated solution 1 (3 nurses, 14 waiting time, 3 overtime), and nondominated solution 4 (4 nurses, 4 waiting time, 0 overtime). In the figure, each patient appointment is represented with a rectangle where the width of the rectangle shows the treatment duration and the height shows the acuity of the treatment. The total acuity assigned to each nurse does not exceed the maximum acuity level. Patients with acuity level 3 cannot be assigned to nurses 3 and 4 whose skill levels are 2.

For nondominated solution 1, the number of patients assigned to nurses 1–3 are 6, 5, and 9, respectively. For nondominated solution 4, the number of patients assigned to nurses 1–4 are 5, 3, 7, and 5, respectively. In nondominated solution 1, nurses are utilized at their maximum capacity for most of the day (total acuity assigned to a nurse is equal to the maximum acuity level). There is also very low slack time,
which might cause problems when there is high variability in treatment durations and nurses have breaks.

When the number of nurses increases, the waiting time and overtime decrease as expected. However, the cost of adding one nurse might be higher than the total waiting time and overtime costs. In that case, the decision maker can use another criterion that combines the cost of an additional nurse with patient waiting time and staff overtime costs to determine the optimal number of nurses. In order to calculate the total cost, we should determine the regular cost ($c_r$) of an additional nurse and overtime ($c_o$) cost of an existing full-time nurse per unit time. We also have to estimate the cost of waiting time ($c_w$) with respect to overtime and regular working time costs. When we know all these cost values, total cost can be calculated as ($c_w \times TWT + c_o \times TOT + c_r \times$ total regular working time). If the decrease in total waiting time cost ($c_w \times \Delta TWT$) and overtime cost ($c_o \times \Delta TOT$) is larger than the increase on regular
cost of an additional nurse ($c_r \times$ total regular working time of an additional nurse), then it will be beneficial to have one more nurse. In our numerical example, if we take $c_w=1$, $c_r=1$, and $c_o=1.5$ to compare nondominated solutions 1 and 4, the decrease in total waiting time and overtime costs will be $(1 \times (14 - 4) + 1.5 \times (3 - 0) = 14.5)$ and the increase in regular cost of an additional nurse will be $(1 \times 16 = 16)$. That means, having three nurses is better than having four nurses in terms of total cost.

**Patient scheduling model:**

For patient scheduling, we solve the same numerical example with 20 patients. We consider 3 nurses with skill levels of 3, 3, and 2, and maximum acuity levels of 6, 5, and 5, respectively. Table 4.4 shows the appointment durations, acuity levels, and primary nurses assigned to each patient for the primary nurse model. We solve the proposed multiobjective optimization model to find the optimal appointment times with the objectives of minimizing total overtime and total excess workload.

<table>
<thead>
<tr>
<th>Patient number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment duration</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Acuity level</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Primary nurse</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient number</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment duration</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Acuity level</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Primary nurse</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 4.4:** Numerical example data for primary nurse model

Table 4.5 shows the objective function values of the nondominated solutions for primary care delivery model. When there is no excess workload allowed ($E_s = 0$), the patient scheduling model gives 1 nondominated solution. When the upper bound on
total excess workload is per slot is increased to 6 \( (E_s = 6) \), the patient scheduling model gives 3 nondominated solutions including the one found by \( E_s = 0 \).

<table>
<thead>
<tr>
<th>Nondominated solution</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper bound on excess ( (E_s) )</td>
<td>0, 6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Obj: Total excess workload</td>
<td>0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Obj: Total overtime (slots)</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total excess workload per slot (min, average, max)</td>
<td>0, 0, 0</td>
<td>0, 0.2, 1</td>
<td>0, 0.4, 2</td>
</tr>
<tr>
<td>Total overtime per nurse (slots) (min, average, max)</td>
<td>0, 0.7, 2</td>
<td>0, 0.3, 1</td>
<td>0, 0, 0</td>
</tr>
</tbody>
</table>

Table 4.5: Nondominated solutions for patient scheduling for primary care delivery model

Figure 4.5 shows the appointment schedule for two nondominated solutions: nondominated solution 1 and nondominated solution 3. Based on the provided nondominated solutions, decision maker can decide whether a part time nurse is required or not. Similar to nurse assignment model, the overtime cost and cost of part-time nurse can be compared. If overtime cost per unit time is \( c_o \), and cost of part-time nurse per unit time is \( c_p \), then it will be more beneficial to use a part-time nurse when total overtime cost \( (c_o \times TOT) \) exceeds the total part-time nurse cost \( (c_p \times \text{total duration part time nurse is required}) \). In our example, if patients 6 and 7 are assigned to a part-time nurse and patient 16 is scheduled to arrive at time slot 7 instead of 4, then a part-time nurse is enough for 4 slots. If patient schedule cannot be changed at this point, then the part time nurse is required at time slots 4, 5, 6, and 12 to share the workload of primary nurse 2.

For this small example, it is easy to find an alternative schedule when one part-time nurse is added to the team. However, as the number of nurses with excess workload increases, it will become more difficult to find a solution manually. In that case, the
Figure 4.5: Gantt chart of (a) Nondominated solution 1 (Maximum total excess workload allowance $E_s$ is 0 (and 6), total excess workload is 0, total overtime is 2), (b) Nondominated solution 3 (Maximum total excess workload allowance $E_s$ is 6, total excess workload is 7, total overtime is 0)

The first constraint of the primary care delivery model (constraint 1.b) can be updated as follows:

$$\sum_{s=H_i}^{S} y_{i,r,s} + \sum_{g \in G_i; s=H_g}^{S} y_{i,g,s} = 1 \quad i = 1 \cdots P$$ (1.c)

where $G_i$ is the set of part-time nurses that can be assigned to patient $i$. This constraint makes sure either the primary nurse or a part-time nurse from set $G_i$ is assigned to patient $i$. The revised model assigns nurses to patients, and finds an optimal appointment schedule. In this study, since our aim is to provide alternative solutions that minimize total excess workload and total overtime simultaneously, this last model that assigns patients to primary or part-time nurses will not be solved in the computational study section.
4.3.4 Spreadsheet-based optimization tools

Our aim is to provide optimization tools that can easily be used by nurse managers and schedulers. We developed spreadsheet-based optimization tools to solve nurse assignment and patient scheduling problems. The optimization tool uses Opensolver to solve the proposed models. Opensolver is an Excel VBA add-in that extends Excel’s built-in Solver capabilities with a more powerful linear programming solver. It is developed and maintained by Andrew Mason and students at the Engineering Science Department, University of Auckland, New Zealand [2, 38].

Figure 4.6 shows the screenshots of the tool for nurse assignment model. The patient information (patient ID, name, appointment time, treatment duration, and acuity level), nurse information (nurse ID, name, skill level, maximum acuity level, shift start time, and shift end time) and clinic hours (start time and end time) are the inputs to the model. After the user enters all the required information in light blue area, and presses the “Solve” button, the optimization model is solved and the solution is displayed in the dark blue area. The solution gives nurse names assigned to each patient, actual treatment start times, waiting times, completion time of last treatment for each nurse, total patient waiting time and total overtime.
<table>
<thead>
<tr>
<th>Nurse ID</th>
<th>Nurse Name</th>
<th>Nurse Skill Level</th>
<th>Max. Acuity Level</th>
<th>Shift Start Time</th>
<th>Shift End Time</th>
<th>Actual End Time</th>
<th>Clinic Start Time</th>
<th>Clinic Official End Time</th>
<th>Total Waiting Time</th>
<th>Total Overtime Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Laney</td>
<td>3</td>
<td>3</td>
<td>8:00</td>
<td>16:00</td>
<td>16:30</td>
<td>8:00</td>
<td>16:00</td>
<td>1:30</td>
<td>0:30</td>
</tr>
<tr>
<td>2</td>
<td>Amy</td>
<td>3</td>
<td>5</td>
<td>8:00</td>
<td>16:00</td>
<td>16:30</td>
<td>8:00</td>
<td>16:00</td>
<td>1:30</td>
<td>0:30</td>
</tr>
<tr>
<td>3</td>
<td>Cherry</td>
<td>2</td>
<td>5</td>
<td>8:00</td>
<td>16:00</td>
<td>16:30</td>
<td>8:00</td>
<td>16:00</td>
<td>1:30</td>
<td>0:30</td>
</tr>
<tr>
<td>4</td>
<td>Linda</td>
<td>2</td>
<td>4</td>
<td>8:00</td>
<td>16:00</td>
<td>16:30</td>
<td>8:00</td>
<td>16:00</td>
<td>1:30</td>
<td>0:30</td>
</tr>
</tbody>
</table>

3. Nurse information is another input to the model. Nurse ID, nurse name, nurse skill level, maximum acuity level, shift start time and shift end time for all nurses on that day should be entered in the light blue area.

4. The regular working hours for the infusion clinic are required for the model. Clinic start time and end time should be entered in the light blue area.

5. Please click on Solve button to solve the optimization model. When the model is solved, the solution will be displayed in the dark blue cells.

For patients, the assigned nurse, treatment start time, and waiting time with respect to appointment time will be displayed.

For nurses, the actual end time with respect to the current nurse-patient assignments will be displayed.

The performance measures including total waiting time and total overtime will be displayed for the optimal nurse assignment.

Figure 4.6: Screenshot of the spreadsheet-based optimization tool for nurse assignment model
Figure 4.7 shows the screenshot of the spreadsheet based tool that solves the patient scheduling problem for primary nurse model. Similar to functional nurse assignment model, the user needs to enter patient information (patient ID, name, assigned nurse ID and treatment duration), nurse information (nurse ID, name, maximum acuity level, shift start and end time) and clinic start and end times in the light blue area. A maximum overtime allowance is required to determine the maximum number of slots required in the proposed model. The upper bound on total excess workload should also be provided by the user. After the model is solved, treatment start times, total overtime, total excess workload, excess workload in each slot for each nurse are provided in the dark blue areas.
4.4 Computational study

We performed a computational study to evaluate the performance of proposed nurse assignment and patient scheduling models. We solve 30 problems with different number of patients, patient acuities, and treatment durations. Each problem represents a single day with a set of patients who have to be treated on the same day. There are 40 to 68 patients per day, each patient has an acuity level from 1 to 3, and their treatments last from 1 to 9 slots (30 minutes to 4.5 hours). Table 4.6 shows the average duration and acuity level per patient, and total workload for each problem. Total workload, which is calculated by multiplying the acuity level and the treatment duration for each patient, shows the total nursing time requirement on a given day. Since the number of nurses affects patient waiting times and nurse overtime, we use different number of nurses changing between 5 and 7. Table 4.7 shows the skill levels and maximum acuity levels of nurses used in the computational studies. All problems are solved using IBM ILOG Cplex 12.0. The computational results are presented for functional care delivery model and primary nurse model separately in the following sections.

4.4.1 Computational results for the functional care delivery model

The multiobjective optimization models are solved using $\epsilon$-constraint approach. The total regular working hours per day is assumed to be 8 hours (16 slots). Since overtime is allowed, a maximum overtime of 8 slots (4 hours) is considered. If the current patient mix cannot be scheduled within 24 slots (16 regular + 8 overtime) with
<table>
<thead>
<tr>
<th>Problem no.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>43</td>
<td>47</td>
<td>45</td>
<td>43</td>
<td>57</td>
<td>48</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>Average duration, $\sum_i D_i/n$</td>
<td>3.09</td>
<td>3.04</td>
<td>3.16</td>
<td>3.37</td>
<td>2.91</td>
<td>3.23</td>
<td>3.21</td>
<td>3.57</td>
</tr>
<tr>
<td>Average of acuity level, $\sum_i A_i/n$</td>
<td>1.6</td>
<td>1.53</td>
<td>1.71</td>
<td>1.63</td>
<td>1.54</td>
<td>1.67</td>
<td>1.67</td>
<td>1.76</td>
</tr>
<tr>
<td>Total workload, $\sum_i (A_i \times D_i)$</td>
<td>245</td>
<td>255</td>
<td>291</td>
<td>294</td>
<td>295</td>
<td>297</td>
<td>298</td>
<td>311</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem no.</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>40</td>
<td>55</td>
<td>49</td>
<td>59</td>
<td>53</td>
<td>53</td>
<td>58</td>
<td>54</td>
</tr>
<tr>
<td>Average duration, $\sum_i D_i/n$</td>
<td>3.78</td>
<td>3.11</td>
<td>3.49</td>
<td>3.24</td>
<td>3.45</td>
<td>3.47</td>
<td>3.38</td>
<td>3.31</td>
</tr>
<tr>
<td>Average of acuity level, $\sum_i A_i/n$</td>
<td>1.8</td>
<td>1.6</td>
<td>1.65</td>
<td>1.66</td>
<td>1.74</td>
<td>1.72</td>
<td>1.69</td>
<td>1.72</td>
</tr>
<tr>
<td>Total workload, $\sum_i (A_i \times D_i)$</td>
<td>321</td>
<td>328</td>
<td>333</td>
<td>348</td>
<td>354</td>
<td>357</td>
<td>359</td>
<td>396</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem no.</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>57</td>
<td>64</td>
<td>53</td>
<td>59</td>
<td>53</td>
<td>62</td>
<td>55</td>
<td>61</td>
</tr>
<tr>
<td>Average duration, $\sum_i D_i/n$</td>
<td>3.56</td>
<td>3.34</td>
<td>3.66</td>
<td>3.39</td>
<td>3.75</td>
<td>3.29</td>
<td>3.73</td>
<td>3.31</td>
</tr>
<tr>
<td>Average of acuity level, $\sum_i A_i/n$</td>
<td>1.61</td>
<td>1.67</td>
<td>1.81</td>
<td>1.66</td>
<td>1.75</td>
<td>1.69</td>
<td>1.87</td>
<td>1.74</td>
</tr>
<tr>
<td>Total workload, $\sum_i (A_i \times D_i)$</td>
<td>397</td>
<td>404</td>
<td>410</td>
<td>411</td>
<td>412</td>
<td>414</td>
<td>416</td>
<td>418</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem no.</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>59</td>
<td>56</td>
<td>68</td>
<td>55</td>
<td>56</td>
<td>54</td>
</tr>
<tr>
<td>Average duration, $\sum_i D_i/n$</td>
<td>3.53</td>
<td>3.66</td>
<td>3.34</td>
<td>3.87</td>
<td>3.96</td>
<td>4.07</td>
</tr>
<tr>
<td>Average of acuity level, $\sum_i A_i/n$</td>
<td>1.81</td>
<td>1.84</td>
<td>1.68</td>
<td>2.02</td>
<td>2.07</td>
<td>2.13</td>
</tr>
<tr>
<td>Total workload, $\sum_i (A_i \times D_i)$</td>
<td>445</td>
<td>453</td>
<td>471</td>
<td>510</td>
<td>522</td>
<td>539</td>
</tr>
</tbody>
</table>

Table 4.6: Summary of characteristics of patient mix for 30 problems

<table>
<thead>
<tr>
<th>Number of nurses</th>
<th>Skill levels</th>
<th>Maximum acuity levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(3, 3, 2, 2, 2)</td>
<td>(6, 5, 5, 4, 6)</td>
</tr>
<tr>
<td>6</td>
<td>(3, 3, 2, 2, 2, 3)</td>
<td>(6, 5, 5, 4, 6, 5)</td>
</tr>
<tr>
<td>7</td>
<td>(3, 3, 2, 2, 2, 3, 3)</td>
<td>(6, 5, 5, 4, 6, 5, 4)</td>
</tr>
</tbody>
</table>

Table 4.7: Nurse skill levels and maximum acuity levels for each nurse

available number of nurses, then the IP model gives infeasible solution. A computation time limit of 600 seconds is used because of the difficulty of solving large size models in short computation times. Figures 4.8 and 4.9 show the total overtime and total waiting time, respectively, for all nondominated solutions found by solving the 30...
problems in 5, 6 and 7 nurse settings. The problems are ranked according to their workload. So problem 1 has the lowest workload and problem 30 has the highest workload. When the workload is low, 5 nurses can find a solution with zero overtime, and less than 40 slots of total waiting time. The total waiting time reduces to less than 10 slots when more nurses are used. As the workload increases, both total overtime and total waiting time increase. The average total overtime over 30 problems is 4, 6, and 6 slots for 5, 6, and 7 nurse settings, respectively. The average total waiting time over 30 problems is 59, 32, and 25 slots for 5, 6, and 7 nurse settings, respectively. Figure 4.10 shows the trade-off between the two objectives for a single problem. The decision maker can choose one of the solutions based on the importance of each objective and the available number of nurses.

Figure 4.8: Total overtime for all nondominated solutions for 30 problems in 5, 6 and 7 nurse settings in functional care delivery model

Table 4.8 shows the minimum, average and maximum computation times, number of nondominated problems, number of infeasible solutions, and range of total waiting time and total overtime for 5, 6, and 7 nurse settings. The computation time is the total computation time to find all nondominated solutions. For example, the total computation time to find 3 nondominated solutions is 5.9 seconds for problem 17 with 6 nurses. The average computation time per problem is less than 30 seconds.
Figure 4.9: Total waiting time for all nondominated solutions for 30 problems in 5, 6 and 7 nurse settings in functional care delivery model.

Figure 4.10: Pareto optimal solutions for problem number 17 with workload 397.

That means, the proposed integer programming model takes almost no time to solve nurse assignment problem compared to manual assignment, which might take 45-60 minutes depending on patient volume. When we look at the maximum computation times, we see that it takes longer time to find all nondominated solutions when more nurses are available. That is because of the increase in number of decision variables and constraints.
<table>
<thead>
<tr>
<th>Nurse Count</th>
<th>Computation times in seconds (min, avg, max)</th>
<th>Number of non-dominated solutions (min, avg, max)</th>
<th>Number of infeasible problems</th>
<th>Range of total waiting time (slots) (min, max)</th>
<th>Range of total overtime (slots) (min, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 nurses</td>
<td>0.4, 25.8, 291.6</td>
<td>1, 2.2, 5</td>
<td>10</td>
<td>31, 98</td>
<td>0, 15</td>
</tr>
<tr>
<td>6 nurses</td>
<td>0.2, 27.2, 315.2</td>
<td>1, 1.8, 5</td>
<td>4</td>
<td>2, 62</td>
<td>0, 18</td>
</tr>
<tr>
<td>7 nurses</td>
<td>0.4, 27.3, 469.8</td>
<td>1, 1.4, 4</td>
<td>2</td>
<td>0, 108</td>
<td>0, 23</td>
</tr>
</tbody>
</table>

Table 4.8: Functional care delivery model: CPU time, number of nondominated solutions, number of infeasible problems, total waiting time and total overtime

The proposed algorithm finds more nondominated solutions when the number of nurses is small. As the number of nurses increases, the number of nondominated solutions decreases due to the overall decrease in overtime and waiting time. The proposed models may become infeasible when the workload is high and the nurse capacity is not enough. In those cases, the models cannot find any nondominated solution for any of the $\epsilon$ values. For example, when number of nurses is 5, 10 out of 30 problems cannot be solved optimally for any $\epsilon$ value. As the number of nurses increases to 7, the number of infeasible problems with no nondominated solution reduces to 2. This shows that more nurses are required for these infeasible problems.

4.4.2 Computational results for the primary care delivery model

For the primary care delivery model, we solve the same 30 problems. The number of patients, acuity levels, and treatment durations are the same as in Table 4.6. However, since primary nurses should be known in advance, a primary nurse is assigned to each patient randomly while making sure that the nurse’s skill level is enough to treat the
The proposed integer programming model is solved with different number of nurses and excess workload allowances per slot ($E_s$). The number of nurses range from 5 to 7 and excess workload allowance in each slot are 0, 6, and 12. When excess workload allowance is 0, the clinic does not have any part-time nurse. When excess workload allowance is 6, then one part-time nurse who has a maximum acuity level of 6 can be used to share the workload of primary nurses. When it is 12, 2 or 3 part-times nurses can be used with maximum acuity levels of 4-6. Similar to functional care delivery model, the number of slots $S$ is taken as 24 (16 slots for regular working hours, 8 slots for overtime). If the current patient mix cannot be scheduled within 24 slots with available number of primary nurses and excess workload allowance, then the IP model gives infeasible solution. A computation time limit of 600 seconds is used to solve each model with different $\epsilon$ values.

Figures 4.11 and 4.12 show the total overtime and total excess workload, respectively, for all nondominated solutions found by solving the 30 problems in 5, 6 and 7 nurse settings with 0, 6, 12 excess workload allowance. The nondominated solutions are divided into three groups. The first group includes the solutions found by all three excess workload allowances (0, 6, 12), the second group includes the solutions found with excess workload allowances of 6 and 12, and the third group includes the solutions that can only be found with excess workload allowance of 12. The number of nondominated solutions and the range of objective function values increase as the total workload increases. When total workload is small, lower total excess workload allowances ($E_s$ is 0 or 6) are enough to find all nondominated solutions. As the workload increases, higher excess workload allowances ($E_s$ is 6 or 12) are necessary to find more solutions. As the number of primary nurses increases from 5 to 7, the total workload allowance of 12 cannot find any additional nondominated solutions over
the allowance of 6. Therefore, the workload can be handled with only one part-time nurse.

Figure 4.13 shows the trade-off between the two objectives for a single problem. When there are 5 primary nurses, same solutions can be found with 6 and 12 excess workload allowance. That means, scheduling one part time nurse gives same results as scheduling two part time nurses, so the optimal number of part time nurses is one. When the excess workload allowance is 0, no nondominated solution can be found with 5 primary nurses. When there are 6 primary nurses, one solution with total overtime of 7 and total excess workload of 0 can be found with zero workload allowance. The same nondominated solution can be found when workload allowance is increased to 6 or 12. Increasing the workload allowance to 6 or 12 adds more nondominated solutions to the set. That means, when there are 6 nurses, and if total overtime of 7 is acceptable, then there is no need to schedule part time nurses. Otherwise, one part time nurse is needed. Figure 4.13 also shows that when there are 7 nurses, there is no need to schedule part time nurses since total overtime and excess workload are 0.
Figure 4.11: Total overtime for all nondominated solutions for 30 problems in (a) 5, (b) 6 and (c) 7 nurse settings in primary care delivery model; solutions are grouped as: solutions found by when 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed, and only when 12 excess workload allowed.
Figure 4.12: Total excess workload for all nondominated solutions for 30 problems in (a) 5, (b) 6 and (c) 7 nurse settings in primary care delivery model; solutions are grouped as: solutions found by when 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed, and only when 12 excess workload allowed.
Figure 4.13: Pareto optimal solutions for problem number 17 with workload 397; solutions are grouped as: solutions found with 0, 6, and 12 excess workload allowed, 6 and 12 excess workload allowed.

Table 4.9 shows the minimum, average, and maximum computation times, number of nondominated solutions, number of infeasible problems, and range of total waiting time and total overtime for 5, 6, and 7 nurse settings and 0, 6, and 12 excess workload allowances in primary care delivery model. When excess workload allowance ($E_s$) is 0, only one solution is found with total excess workload of zero. The average computation times over 30 problems is 1.3 seconds for 5 nurses, less than 1 second for 6 and 7 nurse settings. When excess workload allowance is increased, more nondominated solutions can be found. The average computation times over 30 problems is less than 3 minutes for all problems. The computation time can be as high as 2555 seconds (43 minutes) for a problem. The number of nondominated solutions can reach to 27 solutions for a problem when excess workload allowance is 12. As the number of nurses increases, the number of infeasible problems decreases as expected. The excess workload allowance also reduces the number of infeasible problems.
### Table 4.9: Primary care delivery model: CPU time, number of nondominated solutions, number of infeasible problems, total excess workload and total overtime

<table>
<thead>
<tr>
<th>Number of nurses / excess workload allowance</th>
<th>Computation times (in seconds) (min, avg, max)</th>
<th>Number of nondominated solutions (min, avg, max)</th>
<th>Number of infeasible problems</th>
<th>Range of total excess workload (min, max)</th>
<th>Range of total overtime (slots) (min, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 / 0</td>
<td>0.5, 1.3, 2.1</td>
<td>1, 1, 1</td>
<td>13</td>
<td>0, 0</td>
<td>0, 24</td>
</tr>
<tr>
<td>5 / 6</td>
<td>0.6, 140, 2550</td>
<td>1, 9.1, 18</td>
<td>2</td>
<td>0, 89</td>
<td>0, 26</td>
</tr>
<tr>
<td>5 / 12</td>
<td>0.5, 138, 2446</td>
<td>1, 10.4, 20</td>
<td>0</td>
<td>0, 162</td>
<td>0, 26</td>
</tr>
<tr>
<td>6 / 0</td>
<td>0.3, 0.8, 4.4</td>
<td>1, 1, 1</td>
<td>7</td>
<td>0, 0</td>
<td>0, 19</td>
</tr>
<tr>
<td>6 / 6</td>
<td>0.4, 171, 2256</td>
<td>1, 9.5, 23</td>
<td>0</td>
<td>0, 95</td>
<td>0, 24</td>
</tr>
<tr>
<td>6 / 12</td>
<td>0.3, 164, 2356</td>
<td>1, 9.8, 25</td>
<td>0</td>
<td>0, 124</td>
<td>0, 24</td>
</tr>
<tr>
<td>7 / 0</td>
<td>0.5, 0.8, 1.7</td>
<td>1, 1, 1</td>
<td>4</td>
<td>0, 0</td>
<td>0, 17</td>
</tr>
<tr>
<td>7 / 6</td>
<td>0.4, 93, 1196</td>
<td>1, 7.5, 27</td>
<td>0</td>
<td>0, 73</td>
<td>0, 26</td>
</tr>
<tr>
<td>7 / 12</td>
<td>0.4, 88, 1205</td>
<td>1, 7.5, 27</td>
<td>0</td>
<td>0, 73</td>
<td>0, 26</td>
</tr>
</tbody>
</table>

#### 4.4.3 Managerial insights

The nurse managers, nurses, and schedulers spend significant amount of time for nurse assignment and patient scheduling every day. Besides these two problems that are solved every day, determining the optimal number of nurses is an important problem for clinic managers. The oncology clinics choose a care delivery model considering several factors including the availability of skilled nurses in that area, staffing costs, patient satisfaction, and patient safety. Some of these factors such as nurse shortage cannot be controlled. Staffing cost is second largest cost after the chemotherapy drug costs in the oncology clinics. The proposed models not only provide optimization methods to reduce the time spent for nurse assignment and patient scheduling tasks, but also provide decision making tools to determine the optimal staffing levels. To help healthcare practitioners better manage their resources, we provide some managerial insights related to use of the proposed methods.
1. The proposed nurse assignment model allows clinic managers to evaluate the trade-off between total patient waiting time, total staff overtime, and cost of additional nurses. The clinic managers using functional care delivery model can determine the optimal staffing levels on a given day by first checking the total workload required to treat all patients. For example, in our computational results, the problems with 5 nurses started becoming infeasible when the total workload started exceeding 410. This threshold can be used to determine the minimum number of nurses required. However, the clinic managers should also look at patient waiting times and clinic overtime to adjust this threshold. For example, for problem 17, the total waiting time ranges between 70 and 98 slots, which corresponds to an average waiting time of 1.2 and 1.7 slots (36 and 51 minutes) per patient. If this waiting time is not acceptable, then the clinic managers can reduce the threshold of 410 to a lower value, where the total waiting time and overtime are at acceptable levels.

2. The proposed patient scheduling model provides several nondominated solutions that minimize total overtime and total excess workload for a given set of primary nurses. The trade-off between these objectives allows the clinic managers to determine the optimal schedule for the primary nurses and the number of part-time nurses needed when the workload is high for the primary nurses on a given day. The total excess workload and maximum excess workload at each slot can be used to determine the optimal number of part-time nurses. For example, if the total excess workload is low, then other nurses that have available capacity can help the primary nurses to cover the extra workload. If the total excess workload is high, then the clinic managers might use part-time nurses.
3. The functional care delivery model finds less nondominated solutions compared to the primary care delivery model. It also finds solutions with less total overtime due to the flexibility of assigning patients to any of the nurses. Even though the functional care delivery model requires low computation times for nurse assignment, the clinic still needs a scheduling method that considers the cumulative number of nurses to find a good initial schedule. If the initial schedule does not consider the total nurse capacity, the nurse assignment might cause high patient waiting times and overtime.

4. The primary care delivery model can find more nondominated solutions compared to functional care delivery model. This is due to multiple alternative appointment schedules that can be found by the proposed model. The nurses can manage their own schedules and the schedulers might find it easier to schedule the patients with a single nurse. However, the primary care delivery model requires a method to determine the primary nurse for each patient before the treatment starts. Otherwise, the workload of nurses might have high variability on different days.

4.5 Conclusion

In this study, we considered two different nursing care delivery models used in the oncology clinics. We proposed two optimization models that consider patient acuities, nurse skills, maximum acuity levels, and nurse working hours. For the functional care delivery model, we proposed a multiobjective optimization model to solve nurse assignment problem with the objectives of minimizing total patient waiting time and
total nurse overtime. For the primary care model, we proposed another multiobjective optimization model to find the optimal appointment times with the objective of minimizing total overtime and total excess workload. By allowing excess workload, one can determine the number of part-time nurses needed in the clinic. The proposed models are solved using $\epsilon$-constraint approach to find all nondominated solutions. The decision maker can choose a solution based on the availability of nurses and importance of each objective function. We developed two spreadsheet-based optimization tools that can easily be implemented in the clinics. The tools use VBA to read the patient and nurse information, and Opensolver to solve the proposed optimization models. The tools can easily be used by nurse managers and schedulers for daily scheduling without prior knowledge of VBA and Opensolver.

In this study, our aim is not to compare these two delivery models, but rather provide optimization tools to reduce the time spent for nurse assignment and scheduling tasks, and provide decision making tools for clinic managers to determine optimal staffing levels in clinics that use these two care delivery models. In order to make a fair comparison between these two models, a more comprehensive study that measures several measures including staffing costs, patient satisfaction, and patient safety is required. Table 4.10 shows the advantages and disadvantages of both care delivery models.
### Table 4.10: Advantages and disadvantages of functional and primary care delivery models

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Functional care delivery model</th>
<th>Primary care delivery model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Patient scheduling is less restricted due to availability of more nurses for assignment;</td>
<td>1. Nurses can build stronger relationships with patients, which improves patient satisfaction;</td>
</tr>
<tr>
<td></td>
<td>2. Nurses can have a more balanced workload due to daily nurse-patient assignment according to patient mix;</td>
<td>2. Nurses can become more knowledgeable about patient’s medical history and treatment plan, and can detect small changes in patient’s condition to react on time.</td>
</tr>
<tr>
<td></td>
<td>3. Less nurses are required;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. The functional care delivery model can easily be implemented without changing the scheduling system.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disadvantages</th>
<th>Functional care delivery model</th>
<th>Primary care delivery model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Continuity of care is reduced due to random assignment of nurses to patients.</td>
<td>1. Patient scheduling is more restricted due to daily workload and schedule of primary nurse;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Nurse workload might vary significantly from day-to-day;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. More nurses are required;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. The panel size for each nurse should be determined carefully before the implementation of primary care delivery model.</td>
</tr>
</tbody>
</table>
Chapter 5

Patient scheduling with uncertain treatment durations

Chemotherapy scheduling is to allocate appointment times to patients considering resource availabilities and treatment durations. Registered nurses are responsible for providing the treatments in infusion clinics while patients are seated in infusion chairs during the treatment session. Scheduling patients without considering the uncertainty in treatment durations may cause patient waiting time or nurse idle time in clinic. In this study, we propose a two-stage algorithm to solve chemotherapy scheduling problem with uncertain treatment durations for patients. In the first stage, we propose a mixed integer programming model to determine appointment schedules. In the second stage, we propose an algorithm to assign patients to chairs according to their appointment times and calculate total patient waiting time and nurse idle time.
5.1 Introduction

As mentioned in Chapter 2, due to individual differences between patients such as degree of illness, sensitivity to the treatment or physical condition on the treatment day, the actual treatment durations may be different than the scheduled treatment durations. Also, if the need for premedications or hydration is not included in physician’s notes, a patient may require longer treatment duration compared to scheduled duration [26]. Delaney et al. [16] show the degree of variability in actual durations by treatment regimen in their study.

In a deterministic system, a patient is scheduled with an appointment time based on the deterministic treatment duration. If the scheduler considers both chair and nurse availability while scheduling the patient, then these resources will be available when patient arrives to the clinic. Since patient’s treatment ends at the scheduled end time in a deterministic environment, no waiting time or idle time will occur. However, in a real clinic environment, scheduling patients based on deterministic treatment durations may cause patient waiting time and nurse idle time when realizations of durations are different from the scheduled ones. To solve patient scheduling problem with uncertain treatment durations, we need to determine a schedule for the patients while considering chair requirements, nurse availabilities, patients’ waiting time and nurses’ idle time, which depend on actual treatment durations.

In this study, we consider appointment scheduling problem for chemotherapy patients with uncertain treatment durations in a primary nursing care delivery setting. Our aim is to determine the optimal appointment schedule with the objective of minimizing expected total waiting time and expected idle time while considering nurse and chair availabilities. Due to the difficulty of considering two resources (chairs and
nurses) simultaneously in an optimization model with uncertain treatment durations, we propose a two-stage algorithm to solve the problem. In the first stage, we use a mixed integer programming model to determine patients’ appointment times with the objective of minimizing expectation and variance of number of chairs required. We consider nurse capacity by not scheduling more than one treatment at any slot. In the second stage, we calculate the expected patient waiting time and nurse idle time for the given appointment schedule (determined in stage 1) and all possible discrete realizations of treatment durations.

The contributions of this study are:

1. This is the first study that solves assignment, sequencing and scheduling problem with uncertain treatment durations and multiple resource types in a healthcare setting.

2. In this study, we develop a two-stage algorithm to solve appointment scheduling problem with uncertain treatment durations. In the first stage, the model determines appointment schedules with the consideration of uncertain treatment durations and nurse availability. In the second stage, an algorithm assigns patients to chairs with given schedule to find expected patient waiting time and nurse idle time.

In the remainder of this chapter, the problem is defined with its underlying assumptions, and a two-stage stochastic programming model is proposed in Section 5.2. In Section 5.3, we present a two-stage algorithm to solve scheduling problem with uncertain treatment durations. A numerical example is presented and discussed in the same section. In Section 5.4, we evaluate the performance of proposed two-stage algorithm
by solving problems with different patient mix and treatment duration distributions. Concluding remarks and future research are provided in Section 5.5.

5.2 Problem definition

In this study, we solve appointment scheduling problem with uncertain treatment durations in an outpatient oncology clinic. The notation that will be used throughout the paper can be seen in Table 5.1. We consider a single nurse (i.e. primary nurse) who should treat $N$ patients on a given day. The patients can be assigned to multiple chairs that are monitored by the primary nurse. The day is divided into $T$ time slots with a slot length of $\Delta$ (i.e. 30 minutes). The patients have uncertain treatment durations, but are scheduled to arrive at the beginning of a slot.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>Number of slots ($t = 1, \ldots, T$)</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of patients ($i = 1, \ldots, N$)</td>
</tr>
<tr>
<td>$K$</td>
<td>Number of chairs allocated to a nurse ($k = 1, \ldots, K$)</td>
</tr>
<tr>
<td>$X_{it}$</td>
<td>Binary variable which is equal to 1 if patient $i$'s treatment is scheduled to start at time slot $t$</td>
</tr>
<tr>
<td>$A_i$</td>
<td>Scheduled appointment time of patient $i$ ($A_i = \sum_{t=1}^{T} tX_{it}$)</td>
</tr>
<tr>
<td>$A_{(i)}$</td>
<td>Scheduled appointment time of $(i)^{th}$ patient ($A_{(i-1)} \leq A_{(i)}$, for all $i$)</td>
</tr>
<tr>
<td>$n_{(i)}^\omega$</td>
<td>Nursing time to start the treatment of $(i)^{th}$ patient in scenario $\omega$</td>
</tr>
<tr>
<td>$d_{(i)}^\omega$</td>
<td>Realization of treatment duration for $(i)^{th}$ patient in scenario $\omega$</td>
</tr>
<tr>
<td>$S_{(i)}^\omega k$</td>
<td>Actual treatment start time of $(i)^{th}$ patient on chair $k$ in scenario $\omega$</td>
</tr>
<tr>
<td>$Y_{(i)k}^\omega$</td>
<td>Binary variable which is equal to 1 if $(i)^{th}$ patient is assigned to chair $k$ in scenario $\omega$</td>
</tr>
<tr>
<td>$W_{(i)k}^\omega$</td>
<td>Waiting time of $(i)^{th}$ patient on chair $k$ in scenario $\omega$</td>
</tr>
<tr>
<td>$I_{(i)k}^\omega$</td>
<td>Idle time before $(i)^{th}$ patient on chair $k$ in scenario $\omega$</td>
</tr>
<tr>
<td>$O^\omega$</td>
<td>Overtime for the nurse in scenario $\omega$</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Total number of scenarios</td>
</tr>
</tbody>
</table>

$\Delta$ | Duration of each time slot |
$\bar{d}_i, \underline{d}_i$ | Lower and upper bounds of patient $i$'s treatment duration in minutes |
$F_i(x)$ | Cumulative distribution function for treatment duration of patient $i$ ($F_i(x) = \Pr\{X_i \leq x\}$) |
$J_i$ | Number of uncertain treatment durations of patient $i$ |
$D_{ij}$ | The $j^{th}$ treatment duration setting for patient $i$, $j \in \{1, 2, \ldots, J_i\}$, $D_{i1} < D_{i2} < \ldots < D_{ij_i}$ |
$P_{ij}$ | The probability of the $j^{th}$ treatment duration setting for patient $i$ |
$r_{itu}$ | Probability that patient $i$ requires a chair at time slot $u$ if the patient starts the treatment at time $t$ |
$\alpha_o$ | Weight of objective $o$ ($o = 1, 2$) |
$ER_t$ | Expected number of chairs required at time slot $t$ |
$MaxER$ | Maximum expected number of chairs required throughout the day |
$VarR_t$ | Variance of number of chairs required at time slot $t$ |
$MaxVarR$ | Maximum variance of chair requirement throughout the day |

$Sce^\omega$ | Treatment duration set of all patients in scenario $\omega$, $Sce^\omega = \{d_{(1)}^\omega, d_{(2)}^\omega, \ldots, d_{(N)}^\omega\}$ |
$Prob^\omega$ | Probability that scenario $\omega$ occurs |
$Waiting^\omega$ | Waiting time in scenario $\omega$ |
$I^\omega$ | Idle time in scenario $\omega$ |
$U^\omega$ | Actual completion time of nursing time in scenario $\omega$ |
$C^\omega_k$ | Actual completion time of chemotherapy treatment in chair $k$ in scenario $\omega$ |
$Cmax^\omega$ | Completion time of all treatments in scenario $\omega$ |
$B^\omega$ | Binary variable which is equal to one if all treatments are completed within the clinic session in scenario $\omega$ |
$MaxC$ | Maximum completion time for all scenarios |

$ProbOnTime$ | Probability of completion on time over all scenarios |

Table 5.1: Notation for patient scheduling problem with uncertainty
We propose a two-stage stochastic programming model, which is presented as follows: First-stage model:

\[ \min \ E_\omega(\theta(A, \omega)) \]  
\( \text{st} \quad \sum_{t=1}^{T} X_{it} = 1 \quad i = 1 \cdots N \)  
\[ \sum_{i=1}^{N} X_{it} \leq 1 \quad t = 1 \cdots T \]  
\[ A_i = \sum_{t=1}^{T} X_{it} \times t \quad i = 1 \cdots N \]  
\[ X_{it} \in \{0, 1\} \quad i = 1 \cdots N, \; t = 1 \cdots T \]  
\[ A_i \geq 0 \quad i = 1 \cdots N \]

In the first stage, we determine the scheduled start time \((A_i)\) for each patient with the objective of minimizing the expected waiting time, idle time and overtime \((E_\omega(\theta(A, \omega))\) (SO.1). Constraint (SC.1) makes sure that each patient will be scheduled with an appointment time. Constraint (SC.2) shows a nurse can start at most one treatment in any given slot. Constraint (SC.3) is used to calculate the scheduled start time of patient \(i\). Constraints (SC.4) and (SC.5) are integrality and non-negativity constraints.
Second-stage model:

\[
\min \theta(A, \omega) = \sum_{i=1}^{N} \sum_{k=1}^{K} (W_{(i)k}^\omega + I_{(i)k}^\omega) + O^\omega \\
\text{st} \quad \sum_{k=1}^{K} Y_{(i)k}^\omega = 1 \quad i = 1 \cdots N, \quad \omega = 1 \cdots \Omega \\
\frac{S_{(i)k}^\omega}{M} \leq Y_{(i)k}^\omega \quad i = 1 \cdots N, \quad k = 1 \cdots K, \quad \omega = 1 \cdots \Omega \\
S_{(i)k}^\omega \geq A_{(i)} \times Y_{(i)k}^\omega \quad i = 1 \cdots N, \quad k = 1 \cdots K, \quad \omega = 1 \cdots \Omega \\
S_{(i)k}^\omega \geq S_{(i-1),k'}^\omega + n_{(i-1)}^\omega \times Y_{(i-1)k'}^\omega \quad i = 2 \cdots N, \quad k, k' = 1 \cdots K, \quad \omega = 1 \cdots \Omega \\
S_{(i)k}^\omega \geq S_{(i')k}^\omega + d_{(i')}^\omega \times Y_{(i')k}^\omega \quad i, i' = 1 \cdots N, \quad i' < i \quad i, k = 1 \cdots K, \quad \omega = 1 \cdots \Omega \\
W_{(i)k}^\omega \geq S_{(i)k}^\omega - A_{(i)} \times Y_{(i)k}^\omega \quad i = 1 \cdots N, \quad k = 1 \cdots K, \quad \omega = 1 \cdots \Omega
\]
The second stage model assigns patients to chairs while minimizing total waiting time, idle time and overtime of each scenario (SO.2). Constraint (SC.5) assigns each patient to a chair. Constraint (SC.6) makes sure there is no treatment start time for patient \(i\) if the patient is not assigned to chair \(k\). Constraint (SC.7) shows that actual treatment start should not be earlier than the scheduled start time. Constraint (SC.8) makes sure that the treatment cannot start until the nurse is available. Constraint (SC.9) ensures patient \(i\)'s treatment cannot start until the treatments of previous patients assigned to the same chair are completed. Constraint (SC.10) calculates the waiting time for each patient. Nurse idle time is calculated by constraint (SC.11). Due to the complexity of the problem, this constraint cannot be formulated by linear formulation. Clinic overtime is calculated using constraints (SC.11) and (SC.12). Constraint (SC.13) and (SC.14) are integrality and non-negativity constraints. To
solve the problem, one has to link the scheduled start time \( (A_i) \) found in first stage with the treatment sequence in the second stage. \( (A_{(i-1)} \leq A_{(i)}, \forall i) \)

This two-stage stochastic mixed integer programming model is presumably NP-hard. Number of decision variables are \( N \times T + N \times K \times \Omega \). When there are 8 patients, 16 time slots, 3 chairs with 10000 scenarios, number of decision variables can reach to 240 thousand, which is extremely difficult to solve. It is also difficult to link the treatment sequence with patients’ scheduled start times using integer programming model. Another difficulty is to solve the problem with non-linear constraints. Due to the difficulty of solving the proposed stochastic programming model, we propose a two-stage algorithm where the appointment times are determined in the first stage using a mixed integer programming model, and chair assignments, waiting time, idle time and overtime are determined at the second stage for each possible scenario. The proposed algorithm is explained in the next section.

### 5.3 Proposed two-stage algorithm

We propose a two-stage algorithm to solve chemotherapy patient scheduling problem with uncertain treatment durations. We determine the appointment times in the first stage, and assign patients to chairs to calculate expected waiting time, idle time, and overtime in the second stage. For the proposed algorithm, we made a few simplifying assumptions to reduce the number of scenarios.

We first assume that the treatment durations have a lower bound \( d_i \) and an upper bound \( \bar{d}_i \). Since patients are scheduled to arrive only at the beginning of time slots, we consider discrete treatment durations, which are calculated as \( D_{i1} = \lceil \frac{d_i}{\Delta} \rceil \),
$D_{i,j} = \lceil \frac{d_i}{\Delta} \rceil$, and $D_{ij} = \lceil \frac{d_i}{\Delta} \rceil + j - 1$ for all $j = 2, ..., J_i - 1$. If the distribution functions of treatment durations are known ($F_i(x)$), then the probability of each duration $D_{ij}$ can be calculated as $P_{ij} = F_i(D_{ij} \times \Delta) - F_i(D_{ij-1} \times \Delta)$. When $P_{ij}$ and $D_{ij}$ are known, we can calculate the probability ($r_{itu}$) that a patient $i$ requires a chair at slot time $u$ if patient starts the treatment at time $t$ using the following equations (1) and (2).

\[
\begin{align*}
r_{itu} &= \sum_{j=1}^{J_i} P_{ij} \quad i = 1 \cdots N, \; t = 1 \cdots T - J_i + 1, \\
u &= t \cdots t + D_{i1} - 1 \\
\end{align*}
\] (1)

\[
\begin{align*}
r_{itu} &= \sum_{j=j'}^{J_i} P_{ij} \quad i = 1 \cdots N, \; t = 1 \cdots T - J_i + 1, \; j' = 2 \cdots J_i, \\
u &= t + D_{i1} - 1 + j' \\
\end{align*}
\] (2)

The treatment duration distributions are assumed to be independent of each other for each patient. The number of chairs required by patient $i$ at any given slot is either 0 or 1, which means it has a Bernoulli distribution once treatment start time is determined. The expected number of chairs required and the variance of chair requirements at slot $u$ can be calculated using equations (3) and (4).

\[
\begin{align*}
ER_u &= \sum_{i=1}^{N} \sum_{t=1}^{T} r_{itu} \times X_{it} \quad u = 1 \cdots T \\
\end{align*}
\] (3)

\[
\begin{align*}
Var R_u &= \sum_{i=1}^{N} \sum_{t=1}^{T} r_{itu} \times (1 - r_{itu}) \times X_{it} \quad u = 1 \cdots T \\
\end{align*}
\] (4)
5.3.1 Stage 1: Mixed integer programming model to determine appointment time

We propose a mixed integer programming model with the objective of minimizing the maximum of expectation and variance of number of chairs required. The proposed model determines each patient’s treatment start time.

\[
\min \alpha_1 \text{Max}ER + \alpha_2 \text{MaxVar}R \tag{O}
\]

\[
\text{st} \quad \sum_{t=1}^{T-D_{ij}+1} X_{it} = 1 \quad i = 1 \cdots N \tag{C.1}
\]

\[
\sum_{i=1}^{N} X_{it} \leq 1 \quad t = 1 \cdots T \tag{C.2}
\]

\[
\sum_{i=1}^{N} X_{i1} = 1 \tag{C.3}
\]

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} r_{itu} \times X_{it} = ER_u \quad u = 1 \cdots T \tag{C.4}
\]

\[
\text{Max}ER \geq ER_t \quad t = 1 \cdots T \tag{C.5}
\]

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} r_{itu} \times (1 - r_{itu}) \times X_{it} = \text{Var}R_u \quad u = 1 \cdots T \tag{C.6}
\]
\[ \text{MaxVarR} \geq \text{VarR}_t \quad t = 1 \cdots T \quad (C.7) \]

\[ X_{it} \in \{0, 1\} \quad i = 1 \cdots N, t = 1 \cdots T \quad (C.8) \]

\[ ER_t, \text{VarR}_t, \text{MaxER}, \text{MaxVarR} \geq 0 \quad t = 1 \cdots T \quad (C.9) \]

The proposed model finds scheduled appointment times while minimizing the maximum expected chair requirement in the clinic and the maximum variance of chair requirement. Since proposed model has multiple objectives, we use the weighted sum approach to find multiple schedules by changing the weight of objective functions. Constraint (C.1) ensures that each patient will be scheduled with an appointment time within \( T \) slots. A nurse can start at most one treatment in any given slot, which is guaranteed by constraint (C.2). Constraint (C.3) makes sure that a patient is scheduled to arrive at time slot 1. Constraint (C.4) calculates the expected number of chairs required at each slot. Constraint (C.5) determines the maximum expected number of chairs. Constraint (C.6) calculates the variance of chair requirement at each slot. Constraint (C.7) determines the maximum variance of chair requirement. Constraints (C.8) and (C.9) are integrality and non-negativity constraints.

5.3.2 Stage 2: An algorithm to determine chair assignment and compute the expected idle time and waiting time

We propose an algorithm to calculate the total expected waiting time and idle time for the appointment schedule determined in the first stage. The proposed algorithm
determines chair assignments and actual start times of all patients, which depend on the realizations of treatment durations. The realizations of treatment durations \(d_{ij}^\omega\) and the probability of each scenario \(\text{Prob}^\omega\) are determined by considering all possible discrete treatment durations \(D_{ij}\) and the probabilities of these durations \(P_{ij}\) used in the proposed mixed integer programming model is stage 1. Figure 5.1 shows the generation of all scenarios and calculation of their probabilities.

\[
\text{Sce}^1 = \{D_{(1)1}, D_{(2)1}, D_{(3)1}, \ldots, D_{(N)1}\}, \quad \text{Prob}^1 = P_{(1)1} \times P_{(2)1} \times P_{(3)1} \times \ldots \times P_{(N)1}
\]

\[
\text{Sce}^\omega = \{D_{(1)\omega}, D_{(2)\omega}, D_{(3)\omega}, \ldots, D_{(N)\omega}\}, \quad \text{Prob}^\omega = P_{(1)\omega} \times P_{(2)\omega} \times P_{(3)\omega} \times \ldots \times P_{(N)\omega}
\]

**Figure 5.1:** Scenario generation and probability calculation
The proposed algorithm (Algorithm 1) starts with determining the sequence of patients according to their appointment times determined in the first stage \( A = (A_1, A_2, ..., A_N) \). The patients are sorted according to their appointment times such that \( A(i) \leq A(i+1) \). The nurse starts the treatments of patients based on that sequence. That means, \( A(i) \) denotes the scheduled start time of the \( i^{th} \) patient to be treated. For each scenario, the actual completion time on all chairs \( (C_{k}) \) and completion time of the primary nurse \( (U^\omega) \) are initialized to zero (step 3). These values are updated as patients are assigned to chairs one-by-one. For each patient \( (i) \), the first available chair \( k^* \) is found (step 5), and patient is assigned to that chair (step 6). The actual start time of the patient, which depends on the appointment time, completion time of the previous patient on chair \( k^* \) and the completion time of the nursing time of the previous patient, who might have been assigned to any chair, is calculated in step 7. A patient’s treatment can start only when the nurse is available, the chair is available and he/she arrives for his appointment. Once patient’s actual treatment start time is determined, the total waiting time and idle time for scenario \( \omega \) is updated (steps 8-9). The actual completion time on chair \( k^* \) and completion time of the primary nurse \( (U^\omega) \) are updated before the next patient is considered (steps 10-11). Steps 4-12 are repeated to calculate waiting time and idle time for all patients. After all patients are assigned, the maximum completion time \( (C_{max}^\omega) \) and whether the treatments are completed within the clinic session \( (B^\omega) \) are determined at step 13. After the performance measures for all scenarios are determined, the expected waiting time, idle time, maximum completion time, and probability of completion on time are calculated in steps 15-18. In the next section, a small numerical example is provided to clarify the basic steps of the proposed two-stage algorithm.
Algorithm 1 Chair assignment and calculation of total expected idle time and waiting time

1: Sort the patients according to their appointment times such that $A_{(i)} \leq A_{(i+1)}$ for all $(i)$.
2: for all $\omega = 1$ to $\Omega$
3: Initialize actual completion times of all chairs and completion time of primary nurse ($C_\omega^k = 0$ and $U_\omega = 0$)
4: for all $i = 1$ to $N$
5: Find the first available chair $k^*$ ($k^* = \arg \min_k C_\omega^k$)
6: Assign patient $(i)$ to chair $k^*$
7: Calculate the actual start time of patient $(i)$ ($S_{(i),k^*} = \max\{A_{(i)}, C_\omega^{k^*}, U_\omega + 1\}$)
8: Update total waiting time ($Waiting_\omega = Waiting_\omega + \max\{0, S_{(i),k^*} - A_{(i)}\}$)
9: Update total idle time ($I_\omega = I_\omega + \max\{0, A_{(i)} - \max\{max_k\{S_{(i-1),k} + 1\}, C_{k^*} - 1\}\}$)
10: Update the completion time of chair $k^*$ ($C_\omega^{k^*} = S_{(i),k^*} + d_\omega^{(i)}$)
11: Update the completion time of primary nurse ($U_\omega = S_{(i),k^*}$)
12: end for
13: Calculate completion time of all treatments for scenario $\omega$ ($C_{max_\omega}$) and determine whether the treatments are completed on time before the clinic session ends ($B_\omega = 1$ if $C_{max_\omega} \leq T$, and $B_\omega = 0$ otherwise).
14: end for
15: Calculate expected waiting time, $E[WaitingTime] = \sum_{\omega=1}^{\Omega} Waiting_\omega \times Prob_\omega$
16: Calculate expected idle time, $E[IdleTime] = \sum_{\omega=1}^{\Omega} I_\omega \times Prob_\omega$
17: Calculate maximum completion time, $MaxC = \max_\omega(C_{max_\omega})$
18: Calculate probability of completion on time, $ProbOnTime = \sum_{\omega=1}^{\Omega} Prob_\omega \times B_\omega$

5.3.3 Numerical example

We consider $N=5$ chemotherapy patients that should be scheduled on a given day with $T=10$ slots, and $K=2$ chairs. Table 5.2 shows the discrete treatment durations $D_{ij}$ and the corresponding probabilities $P_{ij}$ for each patient.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{ij}$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$P_{ij}$</td>
<td>0.33 0.34 0.33</td>
<td>0.33 0.34 0.33</td>
<td>0.33 0.34 0.33</td>
<td>0.33 0.34 0.33</td>
<td>0.33 0.34 0.33</td>
</tr>
</tbody>
</table>

Table 5.2: Numerical example data for patients

We first solve the proposed mixed integer programming model of stage 1 to determine
an appointment schedule. Then, we use the proposed algorithm of stage 2 to determine the chair assignments, and calculate the expected waiting time and idle time. Table 5.3 shows the solution found by the proposed algorithm when the objective function weights are (0.5, 0.5). The scheduled treatment start times for patients 1 to 5 are 6, 8, 7, 1, 2, respectively. The maximum expected chair requirement is 2.0 and the maximum variance of chair requirement is 0.44. The weighted average of these values are minimized by the proposed mixed integer programming model.

<table>
<thead>
<tr>
<th>Patient ID, $i$</th>
<th>Scheduled start time, $A_i$</th>
<th>MaxER</th>
<th>MaxVarR</th>
<th>$E[\text{Waitingtime}]$ (slots)</th>
<th>$E[\text{Idletime}]$ (slots)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3, 4, 5</td>
<td>6, 8, 7, 1, 2</td>
<td>2.0</td>
<td>0.44</td>
<td>1.45</td>
<td>1.56</td>
</tr>
</tbody>
</table>

**Table 5.3:** Mixed integer programming model solution for numerical example with $\alpha_1=0.5$, $\alpha_2=0.5$

The proposed algorithm of stage 2 is used to determine the expected waiting time and idle time over all scenarios ($3^5 = 243$ scenarios for the given example). Figure 5.2 shows the chair assignments, waiting times, and idle times for one of the scenarios with treatment durations of $\{4, 5, 3, 2, 2\}$ and probabilities of $\{0.33, 0.34, 0.33, 0.34, 0.34\}$. Patient (1) is assigned to chair 1 and the completion time on chair 1 is updated as $C_{1}^a = 4$. Patient (2) is assigned to the first available chair, which is chair 2, and the completion time on chair 2 is updated as $C_{2}^a = 6$. Both patients (1) and (2) can start their treatments on their appointment times, because primary nurse and chairs are available. Patient (3) is assigned to chair 1, which has earliest completion time. However, the treatment cannot start until time slot 6, because of the scheduled appointment time of 6. This creates an idle time of 1 slot for the nurse. After patient (3) is scheduled, the completion time of chair 1 is updated to $C_{1}^a = 8$. Patient (4) is assigned to chair 2 and the completion time is updated to $C_{2}^a = 8$. Patient (5)'s
scheduled appointment time is 8, however, neither of the chairs is available at that time. So patient (5) needs to wait until slot 9 to start his treatment. We assign patient (5) to chair 1 with updated chair completion time $C_{1} = 10$. After finishing the chair assignment, the weighted average of waiting time and idle time are calculated as $(0.5 \times 1 + 0.5 \times 1)$. The expected waiting time and idle time for this scenario is $1 \times 0.00428 = 0.00428$, where 0.00428 is the probability of the corresponding scenario and 1 is the weighted average of waiting time and idle time. The expected waiting time and expected idle time over all scenarios are calculated as 1.45 slot (43.5 minutes, 8.7 minutes per patient) and 1.56 slots (46.8 minutes), respectively.

Figure 5.2: Chair assignment, waiting time and idle time calculation for a sample scenario

5.4 Computational study

We perform a computational study to evaluate the impact of patient mix with different service time distributions on appointment schedules generated by the proposed method, and on operational performance measures such as patient waiting time and nurse idle time. We consider two treatment duration distributions (uniform and
truncated normal) with different means and variances. Table 5.4 shows the patient types, distribution parameters, means, variances, discrete treatment durations, and corresponding probabilities for each patient type.

<table>
<thead>
<tr>
<th>Patient type</th>
<th>Distribution of treatment duration</th>
<th>Mean</th>
<th>Variance</th>
<th>Discrete treatment durations ( (D_{iw}) )</th>
<th>Probability of each treatment duration ( (P_{iw}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>Uniform (0, 3)</td>
<td>1.5</td>
<td>0.75</td>
<td>1, 2, 3</td>
<td>0.33, 0.34, 0.33</td>
</tr>
<tr>
<td>LH</td>
<td>Uniform (0, 4)</td>
<td>2</td>
<td>1.33</td>
<td>1, 2, 3, 4</td>
<td>0.25, 0.25, 0.25, 0.25</td>
</tr>
<tr>
<td>HL</td>
<td>Uniform (4, 7)</td>
<td>5.5</td>
<td>0.75</td>
<td>5, 6, 7</td>
<td>0.33, 0.34, 0.33</td>
</tr>
<tr>
<td>HH</td>
<td>Uniform (4, 8)</td>
<td>6</td>
<td>1.33</td>
<td>5, 6, 7, 8</td>
<td>0.25, 0.25, 0.25, 0.25</td>
</tr>
<tr>
<td>LL</td>
<td>Truncated Normal (1.5, 0.25, 0, 3)</td>
<td>1.5</td>
<td>0.21</td>
<td>1, 2, 3</td>
<td>0.16, 0.68, 0.16</td>
</tr>
<tr>
<td>LH</td>
<td>Truncated Normal (1.5, 1, 0, 4)</td>
<td>1.4</td>
<td>0.61</td>
<td>1, 2, 3, 4</td>
<td>0.26, 0.41, 0.26, 0.07</td>
</tr>
<tr>
<td>HL</td>
<td>Truncated Normal (6, 0.25, 4, 7)</td>
<td>5.9</td>
<td>0.09</td>
<td>5, 6, 7</td>
<td>0.02, 0.49, 0.49</td>
</tr>
<tr>
<td>HH</td>
<td>Truncated Normal (6, 1, 4, 8)</td>
<td>5.7</td>
<td>0.33</td>
<td>5, 6, 7, 8</td>
<td>0.14, 0.36, 0.36, 0.14</td>
</tr>
</tbody>
</table>

Table 5.4: Patient types (LL: low mean, low variance, LH: low mean, high variance, HL: high mean, low variance, HH: high mean, high variance)

Since the number of chairs allocated to a nurse also affects the patient waiting time and nurse idle time, we consider different number of chairs between 2 to 4. The weights of objectives change between 0 to 1 with 0.1 increments while making sure the sum of the weights is equal to 1 \( (\alpha_1 + \alpha_2 = 1) \). All mixed integer programming problems are solved using IBM ILOG Cplex 12.5 in the first stage. The proposed algorithm in the second stage is coded and run in Julia [1, 9].

We first solve the proposed integer programming model with different weight combinations to find appointment schedules. The impact of patient mix on appointment schedules is discussed in Section 5.4.1. Then, we assign patients to chairs to calculate the performance measures including expected waiting time, idle time, probability of
completion on time, and maximum completion time among all scenarios. The effect of patient mix and number of chairs on performance measures is discussed in Section 5.4.2.

5.4.1 Impact of patient mix on appointment schedules

*Single patient type problems*

For single patient type problems, we consider 4 settings that change according to service time distributions: i) 12 LL patients, ii) 12 LH patients, iii) 8 HL patients, and iv) 8 HH patients. The number of patients that should be scheduled for each patient type is determined according to maximum treatment duration and nurse capacity. For example, the maximum treatment duration for a patient (patient type HH) with high mean and high variance of service time is 8 slots, which means that the latest treatment start time can be slot 9. Since we consider each nurse can start at most one treatment at any slot, the maximum number of HH patients in a day should not exceed 9. When there are 9 HH patients to be treated in a day, we know that there will be one treatment start at each slot from slot 1 to 9 regardless of the weights of objectives. So, we further reduce the number of patients to 8.

Table 5.5 shows the expected total treatment durations, number of unique schedules, range of $MaxER$ and $MaxVarR$ among all schedules, and the total computation times. The expected total treatment duration ranges from 24 slots to 52 slots. The problems with low mean and low variance patients (LL) have lowest expected total treatment duration. When the expected total treatment duration is low, the maximum expected chair requirement ($MaxER$) throughout the day is low. For example,
the maximum expected chair requirement is 2 for LL problem and 5.5 for HH problem. The CPU times presented in Table 5.5 is the total time spent on computing expected idle time and waiting time for all chair settings and unique schedules. On average, it takes less than 3.4 minutes to solve each type of patient mix problem, which is around 30 seconds for each schedule. The computation times to solve the proposed mixed integer programming model in Stage 1 is less than 1 second.

<table>
<thead>
<tr>
<th>Patient mix</th>
<th>Expected total treatment duration (slots)</th>
<th>Number of unique schedules</th>
<th>$MaxER$ (min, max)</th>
<th>$MaxVarR$ (min, max)</th>
<th>Total CPU time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform LL</td>
<td>24</td>
<td>2</td>
<td>2.00, 2.00</td>
<td>0.44, 0.44</td>
<td>25.27</td>
</tr>
<tr>
<td>Uniform LH</td>
<td>30</td>
<td>2</td>
<td>2.50, 2.50</td>
<td>0.63, 0.63</td>
<td>781.40</td>
</tr>
<tr>
<td>Uniform HL</td>
<td>48</td>
<td>2</td>
<td>4.33, 4.33</td>
<td>0.44, 0.44</td>
<td>1.25</td>
</tr>
<tr>
<td>Uniform HH</td>
<td>52</td>
<td>2</td>
<td>5.50, 5.50</td>
<td>0.63, 0.63</td>
<td>3.05</td>
</tr>
<tr>
<td>Truncated Normal LL</td>
<td>24</td>
<td>2</td>
<td>2.00, 2.00</td>
<td>0.27, 0.27</td>
<td>25.23</td>
</tr>
<tr>
<td>Truncated Normal LH</td>
<td>30.48</td>
<td>2</td>
<td>2.14, 2.14</td>
<td>0.48, 0.48</td>
<td>782.03</td>
</tr>
<tr>
<td>Truncated Normal HL</td>
<td>51.76</td>
<td>2</td>
<td>4.49, 4.49</td>
<td>0.27, 0.27</td>
<td>0.82</td>
</tr>
<tr>
<td>Truncated Normal HH</td>
<td>52</td>
<td>3</td>
<td>5.50, 5.50</td>
<td>0.49, 0.49</td>
<td>4.25</td>
</tr>
</tbody>
</table>

Table 5.5: Expected total treatment durations, number of unique schedules, range of $MaxER$ and $MaxVarR$ for single patient type problems

Figures 5.3 and 5.4 show the appointment schedules generated for each patient type setting with uniform and truncated normal service time distributions, respectively. As the mean service time increases, the last slot to schedule a patient reduces as expected. When the mean of service times is low, the last patient is scheduled at slots 13 for high variance, and 14 for low variance. When the mean service time is high, the last patient is scheduled at slots 9-10 for high and low service time variance, respectively.

Two patient type problems
For two patient type problems, we consider 6 settings for each patient type combination: i) 6 LL + 6 LH patients, ii) 5 LL + 5 HL patients, iii) 5 LL + 5 HH patients, iv) 5 LH + 5 HL patients, v) 5 LH + 5 HH patients, and vi) 4 HH + 4 HL patients. Table 5.6 shows the expected total treatment durations, number of unique schedules, range of MaxER and MaxVarR, and computation times. By changing objective function weights, we find more unique schedules compared to single patient type problems. The average time to solve each type of patient mix problem is 4.4 minutes. The uniform distribution gives higher maximum variance of chair requirement compared to truncated normal distribution.

Figures 5.5-5.10 show the appointment schedules generated for each setting and service time distribution. Figure 5.5 shows the appointment schedules for LL+LH patient mix. When the weight of MaxVarR objective is high, the patients with low
Figure 5.4: Appointment schedules for single patient type settings with truncated normal distribution

and high variance are scattered throughout the day (schedules 1-3). As the weight of MaxER objective increases, more patients with high variance are scheduled at later time periods of the session. For LL+HL patient mix setting (Figure 5.6), if the weight of MaxVarR objective is one, the patients with low mean are scheduled at the beginning of the day (schedule 1). As the weight of MaxER increases, the patients with high and low mean are scattered throughout the first 10 slots. After slot 10, the patients with low mean are continued to be scheduled. For LL+HH patient mix setting (Figure 5.7), the patients are scattered throughout the first 9 slots, and patients with low mean and low variance are continued to be scheduled after slot 9 for all weight combinations. In LH+HH setting (Figure 5.8), low mean and high mean patients are scheduled throughout the first 9 slots. The patients with low mean are scheduled after the last HH patient, because LH patients can still finish the treatment.
<table>
<thead>
<tr>
<th>Patient mix</th>
<th>Expected total treatment duration (slots)</th>
<th>Number of unique schedules</th>
<th>$MaxER$ (min, max)</th>
<th>$MaxVarR$ (min, max)</th>
<th>CPU time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform LL+LH</td>
<td>27</td>
<td>8</td>
<td>2.25, 2.33</td>
<td>0.60, 0.63</td>
<td>1214.03</td>
</tr>
<tr>
<td>Uniform LL+HL</td>
<td>40</td>
<td>5</td>
<td>3.00, 3.34</td>
<td>0.44, 0.66</td>
<td>6.59</td>
</tr>
<tr>
<td>Uniform LL+HH</td>
<td>42.5</td>
<td>5</td>
<td>3.67, 4.00</td>
<td>0.60, 0.82</td>
<td>70.63</td>
</tr>
<tr>
<td>Uniform LH+HL</td>
<td>42.5</td>
<td>4</td>
<td>3.67, 4.08</td>
<td>0.47, 0.82</td>
<td>46.90</td>
</tr>
<tr>
<td>Uniform LH+HH</td>
<td>45</td>
<td>6</td>
<td>3.75, 4.50</td>
<td>0.56, 0.81</td>
<td>132.80</td>
</tr>
<tr>
<td>Uniform HH+HL</td>
<td>50</td>
<td>4</td>
<td>5.00, 5.08</td>
<td>0.44, 0.63</td>
<td>3.36</td>
</tr>
<tr>
<td>Truncated LL+LH</td>
<td>27.2</td>
<td>7</td>
<td>2.07, 2.17</td>
<td>0.39, 0.48</td>
<td>1404.22</td>
</tr>
<tr>
<td>Truncated LL+HL</td>
<td>42.4</td>
<td>3</td>
<td>3.33, 3.84</td>
<td>0.27, 0.40</td>
<td>4.20</td>
</tr>
<tr>
<td>Truncated LL+HH</td>
<td>42.5</td>
<td>6</td>
<td>3.84, 3.86</td>
<td>0.39, 0.50</td>
<td>70.73</td>
</tr>
<tr>
<td>Truncated LH+HL</td>
<td>45.1</td>
<td>4</td>
<td>3.74, 4.23</td>
<td>0.44, 0.53</td>
<td>47.43</td>
</tr>
<tr>
<td>Truncated LH+HH</td>
<td>45.2</td>
<td>7</td>
<td>3.74, 4.19</td>
<td>0.47, 0.72</td>
<td>158.99</td>
</tr>
<tr>
<td>Truncated HH+HL</td>
<td>51.9</td>
<td>3</td>
<td>5.47, 5.47</td>
<td>0.37, 0.37</td>
<td>2.60</td>
</tr>
</tbody>
</table>

Table 5.6: Expected total treatment durations, number of unique schedules, range of $MaxER$ and $MaxVarR$ for two patient type problems

in time with a late treatment start time. In HL+HH setting (Figure 5.9), there is one slot with no treatment start in each schedule to make sure chair requirement is balanced. Figure 5.10 shows the schedules for LH+HL setting. Similar to earlier results, patients with low mean and high mean are scattered around without any particular pattern.

** Four patient type problems **

We consider a single setting with 2 LL + 2 LH + 2 HL + 2 HH patients for each
service time distribution. Table 5.7 shows the expected total treatment duration, number of unique schedules, range of $MaxER$ and $MaxVarR$, and computation times. Figure 5.11 shows the appointment schedules. In Figure 5.11.(a), we observe that patients with low mean are mostly scheduled in the second half of the day. When the service times are uniformly distributed, there are more slots without any treatment start in the first half of the day compared to truncated normal distribution due to high service time variability.

<table>
<thead>
<tr>
<th>Patient mix</th>
<th>Expected total treatment durations (slots)</th>
<th>Number of unique schedules</th>
<th>$MaxER$ (min, max)</th>
<th>$MaxVarR$ (min, max)</th>
<th>CPU time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>34</td>
<td>8</td>
<td>3</td>
<td>0.44</td>
<td>6.04</td>
</tr>
<tr>
<td>Truncated Normal</td>
<td>35.02</td>
<td>5</td>
<td>2.93, 3.33</td>
<td>0.27, 0.45</td>
<td>4.01</td>
</tr>
</tbody>
</table>

Table 5.7: Expected total treatment durations, number of unique schedules, range of $MaxER$ and $MaxVarR$ for four patient type problems

Figures 5.12-5.13 show how the number of chair requirements change throughout the day for uniform and truncated normal distributions, respectively. When the objective
### 5.4.2 Impact of patient mix and number of chairs on performance measures

Figures 5.14-5.17 show the expected waiting time, idle time, probability of completion on time, and maximum completion time, respectively, for all patient mix settings. The
expected waiting time decreases and expected idle time increases as the number of chairs increases. The probability of on-time completion increases with the increase in number of chairs as expected. When there are more patients with high service time mean, the probability of on-time completion reduces.

According to the performance measures included in Figures 5.14-5.17, decision maker can determine the optimal number of chairs that should be allocated to the nurse for a given patient mix. For example, for LL single patient type setting with uniform service time distribution, 3 chair setting gives the same waiting time and idle time as the 4 chair setting. So 3 chair setting would be sufficient for this problem. The decision maker can further compare 3 chair with 2 chair setting. If the waiting time in 2 chair setting is acceptable for the decision maker, then allocating 2 chairs to the nurse will be sufficient. In two patient type problems, except LL+LH problems, 2 chairs setting causes very low probability of on time completion. So the optimal
<table>
<thead>
<tr>
<th>Slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LH</td>
<td>HH</td>
<td>HH</td>
<td>HH</td>
<td>HH</td>
<td>HH</td>
</tr>
<tr>
<td>2</td>
<td>HH</td>
<td>LH</td>
<td>HH</td>
<td>HH</td>
<td>LH</td>
<td>HH</td>
</tr>
<tr>
<td>3</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
<td>HH</td>
<td>HH</td>
<td>LH</td>
</tr>
<tr>
<td>4</td>
<td>HH</td>
<td>HH</td>
<td>LH</td>
<td>HH</td>
<td>HH</td>
<td>LH</td>
</tr>
<tr>
<td>5</td>
<td>LH</td>
<td>LH</td>
<td></td>
<td>HH</td>
<td>HH</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>HH</td>
<td>HH</td>
<td></td>
<td>HH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LH</td>
<td>LH</td>
<td></td>
<td>LH</td>
<td>HH</td>
<td>LH</td>
</tr>
<tr>
<td>8</td>
<td>HH</td>
<td>HH</td>
<td>LH</td>
<td>LH</td>
<td>HH</td>
<td>LH</td>
</tr>
<tr>
<td>9</td>
<td>HH</td>
<td>HH</td>
<td></td>
<td>HH</td>
<td>HH</td>
<td>HH</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>LH</td>
<td>HH</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
</tr>
<tr>
<td>11</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
<td>LH</td>
</tr>
</tbody>
</table>

(a)

Figure 5.8: Appointment schedules for two patient type settings (LH+HH) with (a) uniform and (b) truncated normal distribution

number of chairs allocate to the nurse for these problems should be at least 3 chairs. When there are four patient types, the probability of on-time completion is very close to zero for 2 chairs, and above 0.9 for 3 and 4 chairs.
Figure 5.9: Appointment schedules for two patient type settings (HL+HH) with (a) uniform and (b) truncated normal distribution

Figure 5.10: Appointment schedules for two patient type settings (LH+HL) with (a) uniform and (b) truncated normal distribution
Figure 5.11: Appointment schedules for four patient type settings with (a) uniform and (b) truncated normal distribution

Figure 5.12: Number of chairs required for four patient type problem with uniform distribution in (a) schedule 1 and (b) schedule 8
Figure 5.13: Number of chairs required for four patient type problem with truncated normal distribution in (a) schedule 1 and (b) schedule 5
Figure 5.14: Maximum, minimum and average values of (a) expected waiting time and (b) expected idle time for all patient mix and chair settings with uniform service time distribution.
Figure 5.15: Maximum, minimum and average values of (a) probability of completion on time and (b) maximum completion time for all patient mix and chair settings with uniform service time distribution.
Figure 5.16: Maximum, minimum and average values of (a) expected waiting time and (b) expected idle time for all patient mix and chair settings with truncated normal service time distribution.
Figure 5.17: Maximum, minimum and average values of (a) probability of completion on time and (b) maximum completion time for all patient mix and chair settings with truncated normal service time distribution.
5.5 Conclusion

In this study, we solved chemotherapy patient scheduling problem with uncertain treatment durations. We first proposed a two-stage stochastic programming model, which is very difficult to solve due to the large number of scenarios, nonlinear constraints used to calculate idle time, and the need for additional variables and constraints to find a sequence after the appointment times are determined. Therefore, we proposed a two-stage algorithm to determine appointment times, and calculate expected idle time and waiting time. To the best of our knowledge, this is the first study that solves chemotherapy scheduling problem with uncertain service times.

In the first stage of the proposed algorithm, we proposed a mixed-integer programming model to determine an appointment schedule with the objectives of minimizing maximum number of expected chair requirement and maximum variance of chair requirement throughout the day. We used weighted sum method to find different schedules at first stage. In the second stage, we assigned patients to chairs according to the schedule found at stage one and calculated expected waiting time and idle time with all possible realizations. We solved patient scheduling problems with two different uncertain treatment duration distributions.

In this study, we solve the problem in a primary care setting that each nurse is assigned to a set of predetermined patients. One future research direction would be solving the problem in a functional care delivery setting, where a nurse can take care of any patient once available. So, both chair assignment and nurse assignments need to be considered. Our two-stage algorithm does not guarantee optimality with respect to patient waiting time and idle time. Another future research direction is
developing new methods to solve the proposed stochastic programming model to find the optimal appointment schedules.
Chapter 6

Conclusion

In this study, we develop patient scheduling and nurse assignment methods in oncology clinics considering resource availabilities and patients’ treatment durations. We first propose an integer programming model to coordinate oncology consultation and chemotherapy appointments. A simulation model is developed to test the impact of the proposed model. Then we solve nurse assignment problem considering patient acuities in functional care delivery model and patient scheduling problem in primary care delivery model. A spreadsheet based optimization tool is developed for easy implementation. Considering the uncertainty in treatment durations, which is a realistic case in oncology clinics, we develop a two-stage algorithm to solve patient scheduling and chair assignment problem with uncertain treatment durations. In this chapter, we provide concluding remarks in Section 6.1, and some future research directions in Section 6.2.
6.1 Summary

Chemotherapy appointment scheduling and nurse assignment are complex problems that they require schedulers to consider multiple resource availabilities, patients' treatment durations, acuity levels, and appointment types while scheduling patients with appointment times or assigning patients to nurses. Without considering these factors, patients may suffer long waiting times in the clinic and nurses may face unbalanced workload throughout day. To avoid these potential patient and provider issues, we developed advanced patient scheduling and nurse assignment methods with simulation and optimization techniques in Chapters 3, 4 and 5.

In the first part of the study (Chapter 3), we propose an integer programming model to solve patient scheduling problems coordinating oncologist consultation and chemotherapy appointments. Simulation model is used to test the impact of the scheduling method as well as other operational options. According to the study in Chapter 3, using the proposed scheduling method to balance number of appointments (both oncologist consultation and chemotherapy treatment) and number of chairs occupied gives better clinic performance such as patient waiting time and clinic total working time, especially when patient volume is high. Non-staggered nurse schedule, where all nurses start working at the same time every day, lowers the clinic’s total working time. The importance of using an advanced scheduling method increases as the number of available resources decreases and patient volume increases. This part of the study has been published in *International Journal of Production Research* [37].

In the second part of the study (Chapter 4), we proposed two integer programming models to solve nurse assignment problem in functional care delivery model and patient scheduling problem in primary care delivery model, respectively. Patient acuity
levels are used to estimate nursing care requirements and determine nurse staffing level. For easy implementation of proposed models, spreadsheet based optimization tools are developed for both care delivery models. We solved 30 problems with different patient mix alternatives. The average computation time per problem is less than 30 seconds. That means, using the proposed integer programming model can greatly reduce the time spent on nurse assignment and patient scheduling in infusion clinics compared to manually assignment and scheduling. The proposed nurse assignment model also allows clinic managers to evaluate the trade-off between total patient waiting time, total staff overtime and cost of additional nurses. The proposed patient scheduling model provides several nondominated solutions that minimize total overtime and total excess workload for a given set of primary nurses. It allows clinic managers to determine the optimal schedule for the primary nurses and number of part-time nurses needed. It is the first study that solves acuity-based nurse assignment and patient scheduling problems in an outpatient setting. This part of the study has been published in *Health Care Management Science* [36].

In the third part of the study (Chapter 5), we considered uncertain treatment durations when scheduling patient appointments. Without considering the variation of treatment durations, nurse idle time occurs when patient finishes the treatment earlier than the scheduled time, or patient waiting time occurs for the next patient, when treatment finishes later than the scheduled duration. We developed a two-stage algorithm to solve patient scheduling and chair assignment problem with uncertain treatment durations considering multiple resource availabilities and unknown sequence. This algorithm calculates expected waiting time and idle time over all possible realizations of uncertain treatment durations. We solved 11 problems with different patient mix combinations with two different service time distributions (uniform and
truncated normal distribution). We also used different objective function weight alternatives for two objectives to find different schedules. The average computation time for each problem is less than 4 minutes. This is the first study that solves scheduling problem with uncertain treatment durations, multiple resource types, multiple servers and unknown sequence.

6.2 Future research directions

There are a number future research directions to extend this research. In this study, we developed an integer programming model to coordinate oncologist and chemotherapy appointments with the objective of balancing the workload. However, the patients require other resources such as phlebotomists for lab tests and pharmacists for drug preparation. One future research area is to incorporate these additional resources to the optimization model.

We considered patient acuity levels when assigning patients to nurses in functional care delivery model and scheduling patients with an appointment time in primary care delivery model. We used patient acuity levels to estimate nurses’ workload in the study so that the workload throughout the day can be balanced. We solved the problem in a deterministic environment. In functional care delivery model, patients’ appointment times were assumed to be given, cancellations or add-ons were not considered. A future research direction is to develop an online scheduling method that nurse assignment can be adjusted when cancellations or add-ons occur. An online schedule will first determine nurse assignment at the beginning of the day with given schedule. When cancellations or add-ons come up, the assignment will be adjusted to
meet new requirement that objectives stay optimal with minimal change of existing assignment.

When we solve the patient scheduling problem in primary care delivery model, we assumed the primary nurses are known for each patient. However, this model creates high variability in daily workload. Another future research direction can be developing optimization methods to determine panel size of each nurse. The methods developed to determine panel size should consider the workload requirement of each patient (i.e. number of treatment sessions per cycle, treatment time required according to treatment protocol, patient’s individual physical condition), arrival rate of new patients and treatment completion rate of existing patients.

Uncertain treatment durations of chemotherapy patients have been considered in our study to determine patient schedules. Due to the complexity of the problem, we could not solve the proposed stochastic programming model. One future research direction is to develop new methods to solve the proposed model. Another important future research is to consider other uncertainties such as add-ons or cancellations to minimize patient waiting and nurse idle time. We solved the patient scheduling problem with uncertain treatment durations in a primary care delivery model with single nurse and multiple chairs. The same problem can be solved for functional care delivery model. The problem needs to determine which nurse the patient is assigned to and what the scheduled start time for each patient should be so that idle time for the nurses and waiting time for the patients are minimized.
REFERENCES


[34] **Kallen, M., Terrell, J., Lewis-Patterson, P., and Hwang, J.** Improving wait time for chemotherapy in an outpatient clinic at a comprehensive cancer center. *Journal of Oncology Practice* 8, 1 (2012), e1–e7.


Appendix A

Tables of literature review
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Condotta et al. (2014)   | Multi-criteria optimization  | Create a multi-level template schedule to determine appointment start time and nurse assignments. The first model determines the first day of patient treatment. The second model determines appointment start times and nurse assignments | • Minimize average number of waiting day  
• Minimize clash* density  
• Minimize total number of clashes.  
Clash*: more than one activity occurs during one time slot | Four-stage algorithm:  
• Data generation for artificial patients  
• Generate a template schedule: solve integer programming models with multiple objectives optimized lexicographically  
• Produce a running schedule  
• Reschedule |
| Gocgun and Puterman (2014) | Markov decision process      | Determine treatment dates or divert the patients beyond planning horizon with the constraints of  
• Their target date and tolerance day  
• Daily resource capacity  
• Total number of patients diverted or served by overtime each day | • Minimize the cost of diverting patient  
• Minimize treatment delay beyond the tolerance limits | Approximate dynamic programming;  
Heuristics algorithm:  
• Target policy  
• Tolerance policy  
• Capacity policy  
• Earliest policy  
• Latest policy |
| Hahn-Goldberg et al (2013, 2014) | Constraint programming     | Create a dynamic scheduling template to assign treatment times to patients with the constraints of  
• Pharmacy capacity  
• Chair and nurse capacity  
• Nurse maximum patient monitoring number  
• Sequence of stages  
• Wait time between stages | • Minimize completion time | - |
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Sadki et al. (2010b, 2012) | Mixed integer programming | Determine **patient appointment period and oncologist work schedule** with the constraints of  
- Availability of consultation rooms  
- The patient only sees the referee physician  
- Number of delegated patients to the interns  
- Maximum capacity of beds in the afternoon period | • Balance bed load capacity requirement through the week  
• Reduce the violation of the consultation capacity of physicians and the interns | MIP-based heuristic approach:  
• Schedule morning time slots first to balance bedload and schedule as many patient as possible  
• Schedule afternoon time slots in second step  
• Find the solution with local optimization |
| Sadki et al. (2010a) | Mixed integer programming | Determine **patient appointment time (period of the week and week number)** based on oncologists schedule with the constraints of  
- Bed capacity  
- Physician availability  
- Number of patients delegated to the interns | • Balance bed load without changing the schedule of existing patients |                                                                                                 |
| Sadki et al. (2011) | Mixed integer programming | Determine **consultation and chemotherapy start time** of the patients, and **physician schedules** with the constraints of  
- Oncologist availability  
- Bed availability | • Minimize total weighted cost incurred by patient waiting times  
• Minimize makespan | Lagrangian relaxation-based heuristic |
| Santibanez et al. (2012) | Multi-objective integer programming | Determine **patient appointment time** with the constraints of  
- Each patient must be scheduled  
- Nurse capacity  
- Appointment sequence | Minimize the deviations from soft constraints:  
• Patients’ time preferences  
• Pharmacy capacity  
• Balancing workload between nurses and through the day |                                                                                          |

**Table A.2:** Appointment planning and scheduling studies (cont’d)
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Sevinc et al.    | Integer programming           | Determine the **number of patients scheduled** for lab test every day to keep load around target utilization value. Allocate seats to patients on treatment day with the constraints of
• Infusion chair capacity
• Completion of all treatments before the clinic end time | **Objectives**: Keep patient waiting time predictable and to a minimum
• Optimize infusion chair utilization
• Maximize the number of patients scheduled | An adaptive negative feedback scheduling algorithm to determine number of patients scheduled for lab
Heuristic approaches to allocate seats to patients:
• Pick the seat with minimum remaining capacity
• Pick the seat with highest remaining capacity after the allocation of the current patient |
| Shashaani         | Mixed integer programming and simulation | Determine **treatment start time** for patients considering
• Patient preferences
• Nurse schedules
• Start time constraints (i.e. chemotherapy can not be started until patients finish consultation) | **Objectives**: Minimize clinic close time
• Maximize patients’ preferences on nurses and chairs | |
| Tanaka           | Simulation                    | Determine **appointment times and chair assignments** with the constraints of
• Allowable overutilization
• Underutilization | **Objectives**: Minimize average patient wait time | Bin-packing heuristics:
• Horizontal first-fit (HFF) heuristic
• Duration-based first-fit (DB-FF) heuristic
• Duration-based best-fit (DB-BF) heuristic
• Duration-based worst-fit (DB-WF) heuristic |

**Table A.3:** Appointment planning and scheduling studies (cont’d)
Turkcan et al. (2012) | Mixed integer programming | Determine **new patient infusion treatment days** without changing existing patients’ schedule with the constraints of
- Earliest treatment start date
- Total patient acuity level a nurse can be assigned to per day

Determine **patient appointment times, nurse and chair assignments** with the constraints of
- Nurse and chair availability
- Nurse maximum acuity level constraint

Minimize:
- Delay in treatment starts
- Overutilization and underutilization of resource

Minimize:
- Completion time

Rolling horizon methodology:
- Solve the planning problem every \( \Delta \) days for a planning horizon of \( T \) days \( (\Delta \leq T) \)

Divide nurse and chair assignment model into two parts:
- First to solve nurse patient assignment problem
- Second, assign chair to patients.
<table>
<thead>
<tr>
<th>Study</th>
<th>Patient categories (based on)</th>
<th>Processes considered</th>
<th>Other factors considered</th>
<th>Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed et al. (2011)</td>
<td>Chemotherapy treatment durations</td>
<td>Infusion</td>
<td>• Punctual arrivals (all arrivals are assumed to occur at the beginning of the time period) • Add-ons</td>
<td>• Change arrival pattern (number of arrivals per hour) • Change nurse schedule • Add a nurse</td>
</tr>
<tr>
<td>Baesler and Sepulveda (2001)</td>
<td><strong>Patient flow:</strong> 1) Medical oncology patients 2) Ambulatory treatment patients 3) Injection patients 4) Pre-processed patients</td>
<td>Check-in Vitals Blood draw Lab turnover time Pharmacy Consultation Infusion / injection</td>
<td></td>
<td>• Change number of nurses • Change number of chairs • Change lab capacity • Change pharmacy capacity</td>
</tr>
<tr>
<td>Matta and Patterson (2007)</td>
<td><strong>Patient routing:</strong> 1) A clinic visit only 2) A clinic visit and external test visit 3) A clinic visit, external test visit, and an oncology treatment visit 4) A clinic visit and an oncology treatment visits 5) A lab and leave 6) An oncology treatment visit only</td>
<td>Check-in Blood draw External tests Vitals Consultation Infusion</td>
<td>• Patient arrivals: non-homogeneous Poisson process</td>
<td>• Change arrival pattern (smooth arrivals uniformly, smooth arrivals with front-end loading, schedule all external tests prior to 2pm) • Add chairs/beds • Add doctors to different clinics on different days of the week • Add nurses • Change nurse shifts • Create a satellite express testing center for lab and leave patients • Queuing policy (giving higher priority to patients who require more appointments in more facilities)</td>
</tr>
<tr>
<td></td>
<td><strong>Sets of providers (clinics):</strong> Coagulation, hematology oncology, thoracic oncology, gastrointestinal, sickle cell, hematology, fellow patients, surgical oncology, direct from home to treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table A.5:** Studies that consider patient flow
<table>
<thead>
<tr>
<th>Study</th>
<th>Patient categories (based on)</th>
<th>Processes considered</th>
<th>Other factors considered</th>
<th>Scenarios</th>
</tr>
</thead>
</table>
| Santibanez et al. (2009) | **Clinic:** 1) Medical oncology  
2) Radiation oncology  
3) Surgical oncology  
**Appointment type:** 1) New patient (N)  
2) Follow-up (F)  
3) Inter-program consult (C)  
**Resident/student involvement:** 1) Patient is seen by resident/student and oncologist  
2) Patient is seen by oncologist only | Check-in  
Consultation (scheduling of follow-up appointments)  
Check-out | • Unpunctual arrivals  
• Add-ons  
• Late clinic start time | • Clinic start time (On-time, late)  
• Resident/student involvement (yes, no)  
• Appointment order (Mixed, F-C-N, N-C-F, F-N-C)  
• Appointment duration (actual, 15% increase, 30% increase)  
• Scheduling of add-ons (anywhere, at the end of the day)  
• Exam room allocation (dedicated, pooled) |
| Sepulveda et al. (1999) | **Patient flow:** 1) Medical oncology patients  
2) Ambulatory treatment patients  
3) Injection patients  
4) Pre-processed patients | Check-in  
Vitals  
Blood draw  
Lab turnover time  
Pharmacy Consultation Infusion / injection | | • Scheduling alternatives (increase the number of patients scheduled in the morning)  
• Layout (change the location of pharmacy and lab, additional blood draw rooms)  
• New building |
| Shashaani (2011) | **Appointments:** 1) Oncologist appointment prior to chemotherapy appointment  
2) No oncologist appointment prior to chemotherapy appointment  
**Lab tests** (Lab test required, no lab test required)  
**Port** (patients with port, without port)  
**Patient acuity** (determines nursing time) | Registration  
Vitals  
Blood draw  
Lab turnaround time  
Pharmacy Infusion / injection | • Unpunctual arrivals  
• Add-ons  
• Cancellations  
• Nurse workflow and nursing time based on patient acuity | |

**Table A.6:** Studies that consider patient flow (cont’d)
## Table A.7: Studies that consider patient flow (cont’d)

<table>
<thead>
<tr>
<th>Study</th>
<th>Patient categories (based on)</th>
<th>Processes considered</th>
<th>Other factors considered</th>
<th>Scenarios</th>
</tr>
</thead>
</table>
| Tanaka (2011)       | Treatment types: 1) Chemotherapy treatment 2) Non-chemotherapy treatment 3) Chemo and non-chemo treatment 4) Pre-medication only Pre-medicine duration Treatment duration Required lab test: 1) No lab test 2) CBC test required 3) Chem-7 test required | Check-in Blood draw Lab turnaround time Pharmacy Infusion | • Punctual arrivals  
• Deterministic infusion duration  
• Buffer time between appointments  
• Staggered start times for patients assigned to the same nurse  
• Nurse workflow (Order verification, vitals, IV or port setup, drug setup, second vitals, remove IVs) | • Test different appointment schedule rules generated using bin-packing heuristics  
• Determine the time allocated for pre-treatment processes (check-in, blood draw, order verification)  
• Determine the time allocated for preparation and nursing (pharmacy, vitals, assessment, IV or port setup, drug setup) |
| Woodall et al. (2013) | Cancer center area (patient arrivals originating from different clinics): 1) Surgical oncology 2) Oncology 3) Brain tumor 4) Prostate 5) Surgery 6) Direct arrivals to infusion Treatment type: 1) Infusion 2) Injection Patient acuity (which determines treatment times) | Check-in Blood draw Lab turnover time Pharmacy Radiology Infusion / injection | • Nurse scheduling  
• Nurse workflow (Review patient chart, IV set up, finish infusion, injection) | • Change nurse schedules  
• Change number of nurses (add part time nurses)  
• Increase number of chairs |
| Yokouchi et al. (2012) | Cancer types (determines appointment duration distributions): Gastric, colorectal, liver/cholecyst, pancreas, kidney, breast, gynecologic, lung, other cancers | Blood draw Lab turnaround time Pharmacy Infusion | • Nurse workflow (IV access, premedication, infusion setup, finish infusion) | • Change number of nurses when patient arrivals are fixed  
• Change arrival rate when number of nurses is fixed |
<table>
<thead>
<tr>
<th>Study</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Acuity system</th>
</tr>
</thead>
</table>
| Chabot et al. (2005)        | Develop patient acuity system  
Assign patient to nurse for chemotherapy and determine treatment start time | To balance nurse workload, reduce patient waiting time and overtime, increase staff and patient satisfaction | Define patient acuity by  
• Nursing care time                                                                                   |
| Cusack et al. (2004 a, b, c) | Develop and implement patient acuity system to quantify the nursing care requirements  
Determine number of nurses | Identify workload and allocate staffing properly | Define patient acuity by  
• Nursing time  
• Degree of illness  
• Complexity of care needed                                                                 |
| Dobish (2003)               | Develop patient acuity system  
Assign next day patient to rooms | Maximize number of patients scheduled on the day following their appointment with physician | Define patient acuity by  
• The time required to administer the protocol                                                                 |
| Hawley et al. (2009)        | Develop patient acuity system and determine number of nurses by total patient acuity level | Improve patient experience in infusion center and create a balanced workload for nurses | Define patient acuity by  
• Total treatment time  
• Time with patient and/or family members  
• Blood draws  
• Any additional nursing needs |

Table A.8: Studies that develop patient acuity systems for nurse staffing
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodall et al. (2013)</td>
<td>Mixed integer programming</td>
<td>Determine staffing level and nurse schedule</td>
<td>Minimize nursing shortage</td>
<td></td>
</tr>
<tr>
<td>Hertz et al. (2009)</td>
<td>Mixed integer programming with non-linear constraints</td>
<td>Assign patients to nurses in a home healthcare system with the constraints of:</td>
<td>Balance the visit loads, travel loads and case loads</td>
<td>Tabu search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The deviations of visit load, case load and travel load are within an acceptable range</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Each patient is assigned to exactly one nurse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2009)</td>
<td>Mixed integer programming</td>
<td>Determine nurse-to-patient assignment and nurses’ visiting schedule with the constraints of:</td>
<td>Minimize total travelling time of the nurses over the planning horizon</td>
<td>Two phase heuristic algorithm:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Total travelling time and service time for each nurse should not exceed their working hour</td>
<td></td>
<td>• Assign nurses to the patients by K-means clustering algorithm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Route constraint between patients visited on the same day</td>
<td></td>
<td>• Determine the visiting schedules of nurses by heuristic: construction of initial solution, improvement and insertion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Each patient is assigned to exactly one nurse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Visiting interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• At most N nurses can leave the hospital to serve patients assigned to them</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sub-tour elimination: nurses do not travel to patients who will not be visited on the day</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table A.9: Nurse scheduling and assignment studies**
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mullinax et al. (2002)</td>
<td>Integer programming</td>
<td>Create <strong>neonatal acuity system</strong></td>
<td>Determine nursing care needs</td>
<td>Define neonatal acuity system by</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Assign nurses to zones and patients with the constraints of:</td>
<td></td>
<td>• Respiratory status</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Each patient is assigned to exactly one nurse</td>
<td></td>
<td>• Lab work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Each nurse is assigned to one zone</td>
<td></td>
<td>• Feedings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maximum and minimum assigned acuity among non-admit nurse (nurse who does not admit the patient)</td>
<td></td>
<td>• Vital signs, x-rays and weights</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Total amount of acuity assigned to an admit nurse</td>
<td></td>
<td>• Medications</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Balance nurse workload in each zone</td>
<td>• Suction, chest physiotherapy, wound care and dressing changes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Intravenous fluids</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Miscellaneous</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Instability</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Zone-based heuristic to solve nurse-to-patient assignment problem</td>
</tr>
<tr>
<td>Schaus et al. (2009)</td>
<td>Constraint programming</td>
<td><strong>Assign patients to nurses</strong> with the constraints of:</td>
<td>Balance nurse workload in each zone</td>
<td>Two-step approach:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Location constraints</td>
<td></td>
<td>• First assign nurses to zones</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limited number of patients a nurse could be responsible for</td>
<td></td>
<td>• Assign nurses to patients in the zones.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maximum acuity level per nurse</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table A.10:** Nurse scheduling and assignment studies (cont’d)
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Punnakitikas hem et al. (2006, 2008) | Mixed integer programming | Assign patients to nurses with the constraints of:  
• Direct care must be performed in a given period  
• Every patient must be assigned to a nurse  
• Total workload equals to sum of direct care and indirect care  
• Indirect care constraint that allows indirect care to be performed from the beginning of a given time period until the end of shift | Minimize the penalty of excess nurse workload | Use Bender’s decomposition approach and a greedy algorithm to solve the problem  
Built an IT prototype:  
• Shift information entry via Microsoft Excel  
• Data transferring from user personal computer to workstation  
• And optimal assignment display in Microsoft Excel |
| Rosenberger et al. (2004)     | Integer programming    | Assign patients to nurses with the constraint of  
• Every patient is assigned to a nurse | Minimize excess workload on nurses          |                                                                  |
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Performance measures</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Sundaramoorthy et al. (2009, 2010) | Simulation model | Evaluate nurse-to-patient assignment policies | • Total assigned direct care  
• Total unassigned direct care  
• Total direct care  
• Total time spent in non-patient locations  
• The walking time | • Develop simulation model based on data mining  
• Use Classification and Regression Trees (CART) for prediction and classification of nurse status  
• Include service time by Kernel function  
• Evaluate assignment policies including: random, heuristic, total cluster, total stochastic programming |

Table A.12: Nurse scheduling and assignment studies (cont’d)
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Batun et al. (2011)           | Stochastic programming | Determine **number of operating rooms to open, allocate surgeries to operating rooms, determine the sequence in each room and start time of each surgeon** | • Minimize fixed cost of opening operating rooms  
• Minimize cost of overtime  
• Minimize surgeon idle time | • Introduce a set of feasibility constraints  
• Introduce a set of new valid inequalities  
• L-shaped and L-shaped-based branch-and-cut algorithms |
| Daniels and Kouvelis (1995)   | Robust optimization | Determine **job sequence** on a single machine                                | • Minimize total flow time                                                | • Branch-and-bound algorithm  
• Endpoint sum and endpoint product heuristic: order the job by the sum or product of the lower bound and upper bound of uncertain processing time  
• SEPT heuristic: schedule the job with less lower bound and upper bound of processing time before the job with both higher lower bound and upper bound of processing time |
| Daniels and Carrillo (1997)   | Robust optimization | Determine **job sequence** on a single machine                                | • Maximize the likelihood of achieving flow time performance no greater than a target level. | • SEPT heuristic  
• Determine upper bound on the objective value for a given partial schedule  
• β-heuristic  
• Decomposition heuristic |
| Denton and Gupta (2003)       | Stochastic programming | Determine **appointment times** with fixed job sequence                       | • Minimize waiting time  
• Minimize idle time  
• Minimize overtime | • Sequential bounding algorithm  
• Aggregation bounds |
| Denton et al. (2007)          | Stochastic programming | Determine **surgery sequence and surgery start times** for single operating room | • Minimize waiting time  
• Minimize idle time  
• Minimize overtime | • Heuristic 1: Sequence surgeries in order of increasing mean of durations  
• Heuristic 2: Sequence surgeries in order of increasing variance of durations  
• Heuristic 3: Sequence surgeries in order of increasing coefficient of variation of durations |
<table>
<thead>
<tr>
<th>Study</th>
<th>Modeling method</th>
<th>Problem solved</th>
<th>Objectives</th>
<th>Solving approach</th>
</tr>
</thead>
</table>
| Denton et al. (2010)          | Stochastic programming and Robust optimization       | Determine number of operating rooms to open, assign surgical blocks to operating rooms | • Minimize cost of opening operating rooms  
• Minimize cost of overtime                                                                 | • Use symmetry-breaking constraints  
• Determine upper and lower bounds of number of operating rooms                                          |
| Keller and Bayraksan (2009)   | Stochastic programming                              | Determine job start times                                                       | • Minimize expected cost related job start times  
• Minimize penalty cost of exceeding resource capacity                                        | • L-shaped method to decompose the problem into one master problem and several sub problems  
• Various computational enhancements: trust regions, GUB branching, integer cuts on optimality cuts, LP warm starting, approximate master solve |
| Min and Yih (2010)            | Stochastic programming                              | Assign patients to surgical blocks                                              | • Minimize surgical blocks overtime  
• Minimize patient waiting cost                                                               | • Sample average approximation                                                                                      |
| Mittal et al. (2011)          | Robust optimization                                 | Determine scheduled durations for a set of fixed sequence jobs and determine sequence of a set of jobs on a single machine | • Minimize overage cost and underage cost                                                      | • Global balancing heuristic: find the duration that balance overage cost and underage cost  
• Smith’s ordering heuristic: schedule the jobs in the non-decreasing order of (upper bound –lower bound of processing time)/sum of unit overage cost of each job |
| Montemanni (2007)             | Robust optimization                                 | Determine job sequence on a single machine                                       | • Minimize total flow time                                                               | • Introduce valid inequalities                                                                                   |
| Rachuba et al. (2013)         | Robust optimization                                 | Determine surgery date and assign operating room to each patient              | • Minimize patient waiting times  
• Minimize total overtime  
• Minimize number of patients to be deferred to next planning period                              | • Fuzzy sets approach for multi-criteria optimization with a scenario-based setting                                                                 |

Table A.14: Scheduling with uncertainty studies (cont’d)